Methodological Challenges in Treatment Mediation Analysis: Examples from Studies Targeting Psychological Factors in Patients with Musculoskeletal Pain

Gemma Mansell

Doctor of Philosophy

June 2016

Research Institute for Primary Care & Health Sciences,

Keele University

Declaration

This PhD project was nested within the Psychological Workstream of the Spinal Pain Programme, a National Institute for Health Research (NIHR) funded programme of work (Grant code RPPG-0707-10131) obtained at the Research Institute for Primary Care and Health Sciences by Elaine Hay and colleagues.

The idea for this project was devised following discussion of the development of an intervention to target illness perceptions to improve functional outcomes for patients with low back pain. Previous Research Institute work had focused on prognostic studies and evidence had been gathered on what factors influenced patient outcome, but little work had been carried out to investigate whether these same factors could also be effectively targeted during treatment. This PhD project was devised to address this gap in knowledge.

Throughout the course of this PhD project I have developed the ideas in this thesis under the guidance of my supervisors (Jonathan Hill, Daniëlle van der Windt and Chris Main (Chris took over from Kevin Vowles, who supervised me until August 2012)). My supervisors advised on the planning of all the analyses presented and on the writing and structure of the included chapters. I designed the analysis plans, conducted all analyses and wrote all chapters. I received guidance on search strategies from Joanne Jordan and Nadia Corp (Chapter 4), and the critical appraisal of the studies included in Chapter 4 was aided by Jemma Cowen. Elaine Thomas provided statistical guidance on Chapters 5 and 6 and Kelvin Jordan and Anne Smith provided guidance on the Latent Growth Modelling presented in Chapter 7.

All of the data presented in this thesis was collected prior to my appointment at the Research Institute. I was provided with cleaned datasets from Nadine Foster (BeBACK data presented in

Chapter 5), Jonathan Hill (STarT Back data presented in Chapter 6) and Michael Von Korff (Back In Action data presented in Chapter 7). Data from the IMPACT service (Chapter 8) was provided by Julie Ashworth, which did require cleaning by myself prior to use.

Acknowledgements

I would firstly like to thank my supervisors Dr Jonathan Hill, Professor Daniëlle van der Windt and Professor Chris Main for their help, advice and encouragement throughout this process – without them completing this thesis would not have been possible.

I have been well supported throughout my time at Primary Care Sciences by everyone I have been in contact with, but would particularly like to thank Jemma Cowen, Nadia Corp and Jo Jordan for their help with my systematic review (Jemma for her critical appraisal skills and Jo and Nadia for their input on the search strategy); Nadine Foster and Elaine Hay for their comments on presentations that have arisen from this thesis; John McBeth, Christian Mallen and Annette Bishop for conducting my doctoral progression and for their helpful comments at this early stage; Barbara Nicholl, Gwen Wynne-Jones, Rosie Lacey, Ben Saunders and Sarah McLachlan for advice and support as office buddies; and Zoe Mayson and Sara Muller for their friendship, good advice and emotional support (especially to Zoe for proof-reading chapters!)

I have also been lucky enough to have met and had contact with many people outside of the department who have provided helpful advice on different aspects of the thesis: Anne Smith, Curtin University (Australia) for her advice on the latent growth modelling; Richard Emsley, University of Manchester for coming down to Keele to give a workshop on mediation analysis and offering general advice and reassurance; Steve Kamper (George Institute for Global Health, Australia) and Peter Kent (University of Southern Denmark) for their collaboration on a mediation analysis workshop at Forum 2012; and Hopin Lee (University of New South Wales, Australia), a fellow student of mediation analysis for being a sounding board for mediation analysis issues and reminding me I am not alone ☺!

Finally, I would like to thank my lay proof-readers (Kelly and Jarv), and my family, friends and especially Chris for their encouragement and support.

Abstract

Musculoskeletal pain, and particularly low back pain, has been identified as a considerable burden to health and society globally. Psychological factors are known to impact on pain and disability and targeting such factors during treatment has been shown to be effective. Optimising treatment effectiveness requires the identification of key mediators that can be targeted during treatment to lead to an improvement in outcome. However, the study designs and methods of mediation analysis most often applied have been found to be flawed, limiting the interpretability of current treatment mediation literature. The aim of this thesis was to identify, apply, and evaluate the methods and analyses used to carry out treatment mediation analysis in psychological intervention studies of musculoskeletal pain populations in order to improve future mediation studies.

A systematic review identified 10 studies of psychological interventions of musculoskeletal pain populations which had carried out a mediation analysis. Although several treatment mediators were identified, the quality of the analyses were poor with few studies using optimal methods. Secondary analysis of observational data identified several potential treatment mediators. Mediation analysis, performed using recommended methods, subsequently tested factors in two trials previously hypothesised by the study authors as being responsible for the study's success (psychological distress and fear avoidance). A final analysis of diary data investigating the value of measuring key mediators during treatment found that this provided more information than pre-post-treatment scores usually included in intervention studies.

This thesis highlights how a few small changes to the RCT design commonly used to test study effectiveness can allow the inclusion of mediation analysis aiming to identify how a treatment works, leading to more streamlined, focused interventions in the future that more directly target factors most likely to lead to disability improvement.

Contents Page

List of Abbreviations

ACT	Acceptance and Commitment Therapy
AMOS	Analysis of Moment Structures
ANCOVA	Analysis of Co-variance
ANOVA	Analysis of Variance
APT	Active Physical Therapy
BeBACK	Beliefs about Back Pain study
CBT	Cognitive Behavioural Theory/Therapy
CI	Confidence Interval
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CMIN	Chi-Square
CSQ-24	Coping Strategies Questionnaire – 24
DF	Degrees of freedom
DV	Dependent Variable
EFA	Exploratory Factor Analysis
EM	Expectation Maximisation
FAM	Fear Avoidance Model
FIML	Full Information Maximum Likelihood
GFI	Goodness of Fit Index
GP	General Practitioner
HADS	Hospital Anxiety and Depression Scale
IPD	Individual Patient Data
IPQ-R	Illness Perceptions Questionnaire - Revised
IV	Independent Variable
KMO	Kaiser-Meyer-Olkin statistic
LBP	Low Back Pain
LGM	Latent Growth (curve) Modelling
MAR	Missing At Random
MCAR	Missing Completely At Random

MI	Multiple Imputation
ML	Maximum Likelihood
MRC	Medical Research Council
MSK	Musculoskeletal
NFI	Normed Fit Index
NHS	National Health Service
NIHR	National Institute for Health Research
NMAR	Not Missing At Random
NRS	Numeric Rating Scale
OA	Osteoarthritis
OBT	Operant Behavioural Therapy
PCA	Principle Components Analysis
PCLOSE	P value for closeness of fit of RMSEA value
PCS	Pain Catastrophising Scale
PSEQ	Pain Self Efficacy Questionnaire
RA	Rheumatoid Arthritis
RCT	Randomised Controlled Trial
RFT	Relational Frame Theory
RMDQ	Roland-Morris Disability Questionnaire
RMSEA	Root Mean Square Error of Approximation
SCT	Social Cognitive Theory
SD	Standard Deviation
SE	Standard Error
SEM	Structural Equation Modelling
SF-12 PCS	Short Form health survey (12 item) Physical Component Subscale
SIP	Sickness Impact Profile
SPSS PASW	Statistical Package for Social Sciences – Predictive Analytics Software
SRM	Self Regulation Model
SRMR	Standardised Root Mean square Residual
STarT Back	Stratified Targeted Treatment approach
TMD	Temporomandibular Disorder

TSK	Tampa Scale for Kinesiophobia
UK	United Kingdom
US	United States
WAD	Whiplash Associated Disorder

List of Appendices

List of Boxes

List of Figures

List of Tables

Chapter 1: Introduction

This thesis explores the methodological challenges that can be encountered when conducting treatment mediation analysis. This first chapter provides the context of this PhD and explains the background to the project and how the question around mediating factors emerged. It also sets out the definitions of mediation and psychological interventions that will be used throughout this thesis and explains why the investigation of psychological interventions is of particular interest. This chapter helps to focus the reader on the rationale for this project by setting out why research of mediating factors in psychological treatments of back pain and other musculoskeletal (MSK) conditions is important and what work needs to be done to address the current gaps in the literature.

1.1 Defining the Problem

The global prevalence of low back pain (LBP) has been estimated to be 9.4% (95% CI 9.0 to 9.8) (age standardised point prevalence in 2010), with a mean age-standardised point prevalence of 15% (95% CI 14.1 to 16.0) in Western European countries (Hoy et al 2014). LBP has been rated as the greatest contributor to years lived with disability (Hoy et al 2014). A review of the epidemiology of pain conditions identified lifetime prevalence estimates for LBP as being between 51% and 84% (Henschke et al 2015), although the authors describe problems with the definitions used across studies to estimate these prevalence estimates, including variability in the definition of an 'episode' of back pain. In the UK, LBP has been found to affect between 49%-80% of the population at some point in their lifetime, leading to high costs in terms of healthcare, workplace absence and impact on the individual (Maniadakis & Gray 2000). In the UK, MSK conditions are predominantly managed in primary care, with back pain being the most common reason for patients to consult (Jordan et al 2010). A recent UK case-control study comparing people with and

without chronic LBP found that healthcare costs for those with chronic LBP, including general practitioner (GP) visits, outpatient referrals and prescriptions, were almost double the costs for those without chronic LBP (Hong et al 2012). GP consultation costs for MSK consultations were third in 2003, behind costs for respiratory and circulatory system diseases (March et al 2014). Other, less well described costs include the impact on activities of daily living, work and support from family or carers (March et al 2014).

The prognosis of patients with an acute episode of LBP is often good (Pengel et al 2003; Menezes Costa et al 2012). Most patients (68%) return to work within a month and 93% are back at work after six months (Wynne-Jones et al 2014). Indeed, return to work and normal activities as quickly as possible is the recommended course of action for sufferers of acute LBP (Main & Williams 2002). However, prognosis varies substantially for individual patients (Hayden et al 2010) and persisting symptoms have been reported by some patients up to a year after their initial consultation (Hayden et al 2010; Menezes Costa et al 2012). Pain duration is only one of many factors that are important in defining further chronicity; other factors include the impact on function and the intensity of use of healthcare services (Cedraschi et al 1999). Further, the definitions of acute and chronic pain, and whether these terms should be used at all, have been disputed over the years (Cedraschi et al 1999; Von Korff & Miglioretti 2005; Dunn & Croft 2006). The label of “chronic” has negative connotations, both for healthcare practitioners (Cedraschi et al 1999; Barker et al 2009) and for patients, often suggesting incurability and inciting fear (Barker et al 2009). Both “acute” and “chronic” have been found to be often misinterpreted by patients, a fact acknowledged by practitioners (Barker et al 2009). One approach to try and address this has been to employ a prognostic approach to defining chronic pain, and incorporate not only pain duration but also key prognostic variables such as depression and pain severity, to examine whether a person’s condition is likely to improve or worsen (Von Korff & Miglioretti 2005). A risk

score for clinical use was developed to support this approach (Von Korff & Miglioretti 2005), with results being replicated in a UK primary care sample (Dunn et al 2008). The replication study found differences compared to the US sample included in the original study in terms of the proportion of patients assigned to each risk category, which the authors suggest could be due to differences between the healthcare systems or the healthcare settings of the two studies (Dunn et al 2008). Further work on developing this risk score to classify chronic pain found that the risk score was much more predictive of future back pain than pain duration only (Von Korff & Dunn 2008). This prognostic approach also takes into account the uncertainty of risk predictions, by defining chronic pain as "possible" or "probable" (Von Korff & Miglioretti 2005; Von Korff & Dunn 2008). Other studies of long-term trajectories suggest that back pain intensity remains fairly stable over time, depending on patients' baseline pain or short-term trajectory (Dunn et al 2013a), and that while most LBP cases are mild, there is individual variation in terms of whether patients improved or had recurrent pain (Cassidy et al 2005). This again suggests that LBP is not as easy to define as the terms "acute" and "chronic" would suggest.

In summary, MSK pain represents a significant burden for the individual and the wider community. It is a common condition that is likely to fluctuate over time, with sufferers consulting primary care when their pain is most severe. The commonly used terms of "acute" and "chronic" pain have been disputed, with a prognostic approach to LBP that captures the risk of a person developing long-term pain being considered more informative. Identification of prognostic factors that can help to improve the prediction of long-term outcomes of people with MSK pain, or identify those most likely to develop long-term disability, has therefore been the focus of research in recent years.

1.1.1 Keele University Research Institute for Primary Care and Health Sciences

This PhD was developed as part of the psychological workstream of the Research Institute's Spinal Pain Programme, funded by a National Institute for Health Research (NIHR) Programme Grant for Applied Research with the aim of investigating and addressing obstacles to recovery through three separate workstreams (physical, psychological and occupational) to incorporate each element of the biopsychosocial model (see Section 1.2). The psychological workstream aimed to develop better targeting of treatment, so that those who were most in need of psychological treatment would receive appropriate care. Much of the evidence for this work was generated from prognostic studies, including previous Research Institute work, which found that psychological factors were important in predicting LBP outcomes in primary care, including illness perceptions (Foster et al 2008; Foster et al 2010; Campbell et al 2013a), fear-avoidance beliefs (Verkerk et al 2012; Wertli et al 2014a) and catastrophising thoughts (Wertli et al 2014b). Work conducted in the Research Institute also showed that a stratified care approach, where patients consulting in primary care with LBP were assessed for their risk of a poor outcome and treated accordingly (the STarT Back trial), was successful at reducing disability and was also more cost-effective than usual care (Hill et al 2011). The STarT Back trial aimed to not only identify particular groups of patients that needed more specialist care, but also to improve the quality of care by tackling modifiable predictors most likely to lead to an improvement in patient outcomes. The most complex patients, with high scores on measures of physical and psychological factors, received specialised physiotherapy treatment to address these factors. However, the inclusion of psychological factors as targets for treatment was based on their prognostic value, irrespective of the extent to which they were potentially modifiable and therefore appropriate treatment targets. Preliminary analysis of a wide range of psychological factors indicated that illness perceptions in particular appeared to be stable over time, suggesting that they may be difficult to influence and were perhaps not modifiable. This highlighted the difference between predictors

and *mediators*: a predictor is a factor that is associated with outcome regardless of treatment received, while (treatment effect) mediators are factors which change as a result of treatment and therefore contribute to the explanation of outcome. Psychological factors are not only strong predictors but were considered potentially modifiable and were therefore selected as the primary vehicle for the investigation of mediators of treatment outcome. The PhD offered an opportunity to investigate the methodological challenges in explaining outcomes of treatment and help improve the design of interventions designed specifically to identify and target treatment mediators in future studies.

1.2 Biopsychosocial Model of Pain

Individual experiences of pain differ depending on not only physiology but also on previous experience of pain, the context within which it is experienced (Cederaschi & Allaz 2005; Linton & Shaw 2011) and what the person believes the pain to mean (Main & Watson 1999; Turk & Okifuji 2002). The amount of physical damage often does not correlate strongly with the amount of pain or disability reported (Sharp 2001) and it is therefore necessary to move beyond a purely biomedical model of pain to a more evidence-based biopsychosocial model (Vlaeyen & Morley 2005; Gatchel & Turk 2008) which acknowledges the influence of people's response to pain (whether gauged in pain, behaviours, thought processes or emotional responses) and the influence of the person's social environment (Waddell 1987). In the context of LBP, Brunner et al (2013) describe both pain and disability as being multidimensional and part of a biopsychosocial model. It follows that any treatment of pain and disability should encompass not only physical aspects of the complaint but also psychosocial aspects. The biopsychosocial model is therefore important not only for understanding how the condition of LBP develops but also how it is treated (Main & Williams 2002).

It is important to recognise that psychological reactions to pain are normal and expected (Main & Watson 1999; Eccleston et al 2013) and should not to be confused with mental illness (Main & George 2011). The experience of pain is hard to ignore (Linton & Shaw 2011) and we are programmed to respond to it (Williams 2002). Certain behaviours, like rest and avoidance of activities, may be adaptive after an acute injury (Turk & Okifuji 2002), but persistence of pain may lead to these strategies becoming maladaptive (Turk & Okifuji 2002; Linton & Shaw 2011). Similarly, attention to pain is adaptive initially as it helps to motivate us to do something about it (Linton & Shaw 2011), but hypervigilance to pain, especially when little can be done to reduce its intensity, may cause additional anxiety and fear (Turk & Okifuji 2002) and lead to further disability (Crombez et al 2012). Psychological factors or experiences related to pain are therefore not pathological, but do sometimes need modification in order for the patient to regain function and the ability to do most of his or her everyday activities.

However, the need to accept the biopsychosocial model is not just limited to clinicians and researchers – patients also need to accept that psychosocial factors play an important role in their pain experience. In a qualitative study of patients with persistent pain (*n*=10), Snelgrove et al (2013) found that most of their participants viewed their pain in a biomedical context, and were wary of 'alternative' treatments that focused on psychosocial aspects of their pain experience. The authors suggested that this focus on physical pain and trying to 'fix' it maintained their pain experiences over time. Pincus et al (2013) also assert that the biopsychosocial model is currently not being well applied in practice, especially in a primary care context, with only one or two of the components being integrated into treatment. It could be that those working in primary care are not always sufficiently skilled to identify psychological factors contributing to the persistence or development of pain-related disability. This could help explain why current primary care psychological treatments do not show large effects. Nonetheless, there is evidence to support psychologically orientated pain management in secondary and tertiary care (Eccleston et al 2009).

1.3 Evidence for the Importance of Psychological Factors in Predicting LBP Outcomes

The majority of evidence for the importance of psychological factors comes from prognostic studies, where psychological factors have been found to predict patient outcome regardless of what treatment a patient received (Hill & Fritz 2011). A number of reviews exist that have investigated the relationships between psychological factors and functional limitations. Factors investigated have included fear-avoidance behaviours and beliefs (Linton 2000; Pincus et al 2002; Geisser et al 2003; Pincus et al 2006; Lakke et al 2009; Ramond et al 2011; Verkerk et al 2012; Zale et al 2013; Wertli et al 2014a), catastrophising thoughts (Geisser et al 2003; Wertli et al 2014b), self-efficacy (Geisser et al 2003; Jackson et al 2014), depression (Linton 2000; Geisser et al 2003; Mallen et al 2007; Ramond et al 2011), psychological distress (Hayden et al 2009), negative cognitions (Hayden et al 2009), recovery expectations (Laisné et al 2012), somatisation (Mallen et al 2007; Laisné et al 2012), passive or maladaptive coping (Linton 2000; Chou & Shekelle 2010; Ramond et al 2011; Laisné et al 2012) and anxiety (Mallen et al 2007). Almost all of these reviews report evidence for the importance of psychological factors, with strong associations with functional limitations both cross-sectionally and over time. However, several reviews also highlight the poor quality of the included studies, stating that the evidence is conflicting and that the heterogeneity across the studies makes it difficult to draw strong conclusions. One study that looked to create new prognostic models for the outcomes of pain, activity limitation and participation restriction (Kent & Keating 2008) found that the conflicting evidence made this task impossible, and concluded that future prognostic studies with better designs and stronger methodology were needed. For fear-avoidance in particular, the evidence is somewhat mixed with some reviews concluding that this factor is important (Linton 2000; Geisser et al 2003; Ramond et al 2011), others reporting a lack of or little evidence for the importance of fear-avoidance (Pincus et al 2002; Pincus et al 2006; Lakke et al 2009), and others still reporting

inconclusive findings for this factor (Verkerk et al 2012; Wertli et al 2014). In primary care studies specifically, there is inconsistent (and insufficient) evidence for fear-avoidance beliefs (Verkerk et al 2012; Wertli et al 2014a) and catastrophising thoughts (Wertli et al 2014b) as predictors of LBP outcomes.

Overall, this body of research highlights the importance of psychological factors as predictors of LBP outcomes and evidence for their potential importance in influencing treatment. Guidelines on the management of LBP also recommend screening for and management of psychological factors (Koes et al 2010; Mayer et al 2010) although actual formal treatment of psychological factors seems to remain within the remit of secondary care according to some guidelines (Mayer et al 2010). Psychological treatments such as cognitive-behavioural therapy (CBT) for chronic LBP is also recommended (Eccleston et al 2009; NICE Low back pain guideline 2009 (CG88)), although it is noted that to treat a complex chronic pain condition a range of interventions may be required (Airaksinen et al 2006).

However, it is still unclear how psychological factors can best be utilised to improve patient outcomes (Linton & Shaw 2011). Some studies have carried out process analyses, which have investigated whether change in psychological variables is related to change in outcome. These studies move on a step from prediction studies by looking at how variables relate to each other over time and within a specific treatment. For example, Vowles et al (2007) found that changes in measures of catastrophising thoughts and acceptance contributed to changes in the outcomes of depression and physical performance in patients with chronic pain who had completed CBT treatment. This finding was supported by Baranoff et al (2013) who found statistically significant correlations between pre-post treatment change in catastrophising and acceptance as predictors of pre-treatment to three-month follow-up change in disability, a finding which was still present at six-month follow-up (Baranoff et al 2014). Other studies have gone further by including

assessments made during treatment, or at multiple time points, in their analysis. Burns et al (2003) investigated whether catastrophising thoughts and helplessness were associated with outcomes of CBT in 90 patients with chronic pain. Assessments were made at pre-, mid- and post-treatment. They found that early (pre- to mid-treatment) changes in both catastrophising thoughts and helplessness were associated with later (mid- and post-treatment) changes in pain interference after controlling for depression. Bergbom et al (2012) explored changes in depression, catastrophising thoughts, worry, fear and avoidance beliefs on disability in 64 patients with a MSK problem (mostly LBP). Measurements were taken weekly from two weeks before to two weeks after psychological treatment which aimed to address worry, catastrophising thoughts and depression, and early change (pre-treatment to third week of treatment) in all variables except for catastrophising thoughts were associated with outcome. Overall, this evidence suggests that psychological factors are modifiable when targeted within treatments and that change in those factors is related to change in outcome. However, more work is needed in order to investigate how psychological factors can best be utilised in interventions to help improve outcomes in primary care LBP populations.

1.3.1 Psychological interventions in LBP

Interventions designed to influence psychological factors in order to improve functional outcomes in LBP have given mixed results, with some reporting positive findings (Lemstra & Olszynski 2005; Lamb et al 2010) and others not being particularly successful (Jellema et al 2005; Hay et al 2005). The reasons behind this have been explored, with one review identifying key issues which could explain the small treatment effects seen: the heterogeneity of patient populations, treatment content and delivery; issues around suboptimal assessment of back pain; and not enough assessment points being included in trials to show the course of back pain and how this changes over time (van der Windt et al 2008). A recent systematic review of psychological interventions for acute and subacute LBP found that the included studies had either no or inconsistent impact

on outcomes such as pain and functioning (Ramond-Roquin et al 2014). However, the review authors used a very wide definition of "psychosocial interventions" that mostly included an information intervention (reassurance, advice, encouraging patient to stay active). It could be argued that information giving is not a formal "psychological" intervention and that this term should be reserved for interventions that explicitly focus on psychological factors and attempt to change them, such as CBT. It is acknowledged that even within CBT interventions can vary enormously, from specialist care by a clinical psychologist in cases where pain results in a severe psychological impact, to psychologically informed physiotherapy delivered by non-psychologists trained to manage psychological obstacles to recovery (Main & George 2011). In this thesis, "psychological interventions" are defined as those which specifically target psychological factors for change, regardless of the setting, delivery or 'strength' of the intervention.

1.4 The Use of Mediation Analysis to Extend Investigation into the Importance of Psychological Factors in LBP Outcomes

1.4.1 Definition of mediation

In order to improve the effectiveness of future interventions, variables that lead to change must be identified (Smeets et al 2006). Such information can help to tailor treatment in order to most effectively target the key factors that will improve patient outcome (Keefe et al 2004). These factors are known as mediators, and from a psychological treatment perspective are defined as factors which help explain how a treatment works by measuring the change that happens during treatment as well as its subsequent association with outcome (Vlaeyen & Morley 2005). A simple mediation model (involving only one potential mediator) is shown in Figure 1.1 below. Mediation analysis is important not only in terms of showing how a treatment may work, but also in terms of

showing which variables included in an intervention were key to changing the outcomes and suggesting how future interventions can be improved (MacKinnon et al 2007).

Figure 1.1 Example of a simple mediation model

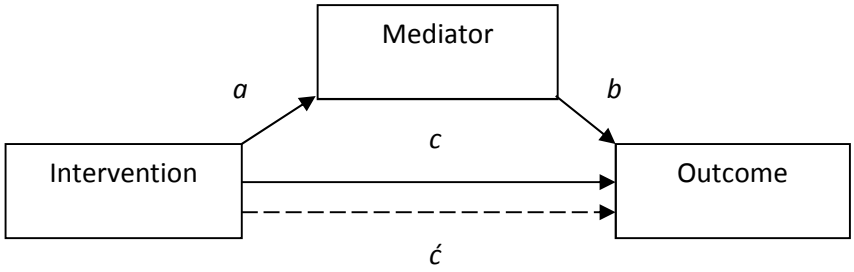


Figure 1.1 presents a causal pathway between the intervention, mediator and outcome. The letters *a*, *b* and *c* denote the direct effects of each variable on the other. The *c* path is the direct effect of treatment on outcome, before taking into account the effects of specific mediating variables. The mediating effect is defined in different ways (see Chapter 3) but the method used in this thesis defines mediation as the product of paths *a* and *b* (*a* x *b*, or *ab*). The *ć* path denotes the total effect of the whole model (*ab*+*c*).

In order for a variable to be considered a potential mediator, it must be potentially modifiable, be correlated with change in the outcome measure (Kazdin 2007) and the intervention should also be associated with change in the mediator (Kazdin 2007). Kraemer et al (2002) also specify that to be a mediator the variable must change during treatment. It has been acknowledged that just because a factor is prognostic (has been found to be related to outcome) does not mean it is necessarily *causally* related to outcome (van der Windt et al 2008). Causality is a key issue in mediation analysis and is explored in depth in Chapter 3.

1.4.2 Related terms and common misinterpretations

Analysis and ideas around mediation have been developed in different fields, leading to different terms being used to describe the same concept. “Intermediate variable” and “intervening variable” are commonly used terms for mediating factors (Kraemer et al 2001). The word mechanism is also often associated with mediation analysis and the two words are often used interchangeably. However, identifying mediators is only a first stage in establishing mechanisms; many mediators may not actually be the key factor responsible for the treatment’s effect despite being causally associated with both treatment and outcome (Laurenceau et al 2007). Another term that is often confused with mediation is moderation, which concerns factors *predicting* (rather than explaining) the effect of treatment. Vlaeyen & Morley (2005) define moderation as a factor that affects how strong a relationship is between an indicator and the outcome and specifies *when and at what* point an effect occurs (i.e. what works for which patient(s)). They define mediation as a factor that helps explain *how* an effect occurs (Vlaeyen & Morley 2005). In the context of treatment, moderators can be used to help identify which individuals or groups of patients are most likely to respond well to a specific treatment, and mediators can be used to understand why a specific treatment is effective, aiming to optimise treatment effects. The terms mediation and moderation are often described together in the literature and it is easy to see that investigation of both of these elements of an intervention would be important and potentially linked, which may be the reason behind the terms being used interchangeably in early literature. Mediation is also often confused with confounding, which occurs when a variable not specified in the causal pathway partly explains the relationship between two variables (MacKinnon et al 2000).

“Indirect effect”, a term from the epidemiological and statistical literature, is also used to describe mediating effects. It has been suggested that a distinction can be made between mediating and indirect effects, as an indirect effect occurs when there is no direct effect of the treatment on

outcome but there may still be an association along a causal pathway, while a mediating effect occurs when there is an effect of the intervention on outcome (Hayes 2009). These terms are often used interchangeably but this suggests that they are not synonymous. Further, Kline (2015) suggests that "mediation" should only be used in a design with time precedence or other aspects that provide support for a causal link, while "indirect effect" can be used in cross-sectional designs. The term 'mediating' effect should therefore perhaps only be used when examining mediators in the context of a specific treatment, where an intervention is used in the study and the effect of this intervention is positive, and that the term 'indirect effect' could possibly be used for mediation analysis in observational designs or for intervention studies in a non-randomised context. In this thesis, the words "mediation" and "indirect effect" are both used depending on the analysis being applied and the evidence being described (i.e. if an author has used the term mediation rather than indirect effect then this term was used when summarising the findings). However, when relating to the empirical work generated as part of the thesis, the term "mediation" will be used to reflect the treatment-specific context of this work.

1.4.3 Action theory and conceptual theory

Chen (1990; 2015) discussed how mediation models of intervention studies can be evaluated by being split into two parts; the action theory (*a* path) and the conceptual theory (*b* path). Action theory is described as the intervention's power to detect the potential mediator while conceptual theory is described as the potential mediator's power to detect the outcome (Chen 2015). It has been suggested that these two elements form the theoretical basis between the two paths (Stanton et al 2013). If the association between the intervention and the potential mediator is weak, this suggests that the action theory has failed – that the intervention is not doing enough to affect the mediator (Chen 2015). If the association between the potential mediator and the outcome is weak, then the conceptual theory has failed – the intervention is targeting the wrong factors for change (Chen 2015), meaning that the underlying theory is wrong. It is not always clear

whether the point of this is to test the theory underlying the intervention or the intervention itself (Coryn et al 2011), but from a clinical perspective, conceptualising a mediation analysis in this way may help to see which path needs strengthening to help improve future studies.

1.5 Psychological Models

Theoretical models can be said to have three main purposes (McCracken & Morley 2014): they can help us integrate our research findings into key principles; they can include specific goals and assumptions; and they can help promote progress through testing and revision. In the field of psychology, and specifically pain psychology, many models have been developed to try and explain the reasons why people may develop persistent pain or other poor outcomes from a pain experience. Painter et al (2008) highlight the importance of theoretical models in explaining how behaviour affects outcome, but their systematic review of the use of theoretical models in health behaviour research showed that only around a third of articles acknowledged a specific theoretical basis. The authors also noted that non-intervention studies were more likely than intervention studies to attempt to apply or test theoretical models. Interventions that are explicitly based on a theory rather than just using it to inform their intervention allow the testing of causal pathways that may underlie the treatment (Michie et al 2008), allowing assumptions about how the theory works to be tested, and also help to generate hypotheses of how we might go about changing those factors or determinants during treatment. Interventions that target these determinants are more likely to be effective (Michie et al 2008). However, few theoretical models have a clear explanation of how they may be operationalised (Michie et al 2008), which leads to theory being used to explain behaviour but not necessarily to suggest how to alter it.

However, not all studies have found the inclusion of theory to lead to more effective interventions. Prestwich et al (2014), in a review of interventions to improve physical activity and diet, found that tailoring an intervention or referring to a specific theory (the Trans-Theoretical Model and Social Cognitive Theory) did not lead to larger intervention effects, although interventions based on a single theory were found to be more effective than those based on multiple theories. The authors did note a number of limitations in their review, including the restrictive inclusion criteria (topic and theory) and problems with treatment fidelity, meaning that interventions may not have been delivered in the optimum way. They suggest that further investigation into specific causal pathways between theory components would be useful.

In the context of mediation, a theoretical basis to the analysis has been suggested to be important in a number of ways. It helps to generate hypotheses about which variables to test as potential mediators, and how they might relate to each other (Kazdin & Nock 2003; Maric et al 2012), leading to stronger models and a stronger justification for any mediating effects found (Mathieu et al 2008). It also allows decisions to be made about when change in the mediator and outcome variables is likely to occur, so that assessments can be made at these points in order to capture this change (Kazdin 2007). Basing mediation analysis on a specific theory has therefore been recommended as an important pre-requisite in mediation analysis in a number of fields (Kazdin & Nock 2003; Johansson & Høglend 2007; Kazdin 2007; Maric et al 2012).

Psychological models of pain imply a person-centred approach, and posit that each individual needs to be treated according to their own personal situation (Linton & Shaw 2011). While the summaries below are not an exhaustive account of the models that have been proposed in the psychological literature, these have been reviewed by Linton & Shaw (2011) who stated them to be the current key models of pain and disability; some of which being specific models of chronic pain (fear-avoidance model (FAM), acceptance and commitment model, misdirected problem

solving model) and others being more generic models of health and pain (self-efficacy model, diathesis-stress model). The FAM is the most researched of the models summarised, especially in the context of mediation, therefore more attention is given to this model.

1.5.1 Fear-avoidance model (FAM)

The term "fear-avoidance" encompasses how both cognitive and behavioural factors play an important role in pain and disability (Boersma et al 2004) by linking pain avoidance as an aspect of pain behaviour (Fordyce 1976; Main et al 2015) with beliefs about pain. The FAM can therefore be described as a specific application of cognitive-behavioural theory (CBT) (Skinner et al 2012). Several FAMs have been proposed, including Lethem et al's (1983) model of pain perception in relation to fear-avoidance and a cognitive model of avoidance proposed by Philips (1987). However, within the pain literature the FAM proposed by Vlaeyen et al (1995a) has received the most attention and it is this version that will be focused on in this thesis. This model states that when an individual experiences pain, a number of thought processes and behaviours then impact on outcome. Negative thought processes such as pain catastrophising (an exaggerated negative perception of a pain experience (Turk & Wilson 2010)) lead to fear and avoidance of behaviours that the patient feel may cause damage or further pain, and may be enhanced further by hypervigilance to perceived threat-related stimuli, which in turn lead to disuse, further disability and depression (Vlaeyen & Linton 2000). This becomes a cycle in which further experiences of pain lead to further catastrophising thoughts and fear-avoidance behaviour, exacerbating the problem (Vlaeyen & Linton 2000). In contrast, individuals who do not catastrophise and/or do not avoid certain behaviours or tasks in an attempt to avoid pain are more likely to recover. Adaptations to Vlaeyen et al's model have been proposed (e.g. Norton & Asmundson's (2003) extension to include an explanation of how physiological processes might be involved in fear-avoidance behaviour).

The evidence for the main components of the FAM (fear-avoidance beliefs and behaviour, catastrophising thoughts, hypervigilance and depression) indicates that each of these components are important in predicting pain and disability outcomes. Reviews of evidence for specific components of the FAM have concluded that there is a positive relationship between the model components and disability (Turk & Okifuji 2002; Leeuw et al 2007a; den Hollander et al 2010), and that fear-avoidance beliefs and behaviour can lead to poor physical function (Vlaeyen & Linton 2000; Wertli et al 2014a).

However, not all of the evidence for the predictive value of the FAM components is supportive. While there is strong cross-sectional evidence that supports the FAM, the evidence in prospective studies is mixed (Pincus et al 2010). Also, work using structural equation modelling (SEM), a technique which builds on regression analysis and allows for paths between variables to be investigated, has found that pain-related fear is not necessarily associated with pain intensity in patients with pain, thereby casting doubt on the cyclical hypothesis of the FAM (Gheldof et al 2010), and suggesting a more direct link between catastrophising thoughts and depression than the model hypothesises (Cook et al 2006). The evidence includes intervention studies where fear-avoidance was not necessarily targeted for change, and the authors of these reviews are cautious about stating whether reducing fear-avoidance and catastrophising thoughts directly leads to improved outcome (Vlaeyen & Linton 2000; Wertli et al 2014a). Finally, partial support has been found for catastrophising thoughts as a mediator between fear-avoidance beliefs and both pain and disability (George et al 2011) but this was a cross-sectional study so the hypothesised order of the FAM variables cannot be tested. A process analysis study (Wideman et al 2009) which hypothesised that early changes in catastrophising thoughts would be associated with late changes in fear-avoidance behaviour and depression, which would then predict return to work outcome, failed to show a positive relationship between these variables despite each being

individually associated with the outcome. The authors suggested that the lack of evidence for a temporal link between catastrophising thoughts and fear-avoidance behaviour could be because the variables were not measured at optimal time points, as the FAM gives no guidance on when these variables are likely to change.

In addition, intervention studies that have been designed to target fear-avoidance behaviour have not shown large improvements in outcome (Pincus et al 2010). This may be due to the construct of fear-avoidance behaviour, and also the patients most at risk of being fear-avoidant, being poorly identified in interventions (Pincus et al 2010). Fear-avoidance can be related to either behaviour or beliefs in different contexts, such as damage caused by movement, particular activities, or both and current measures of fear-avoidance do not distinguish these different types clearly (Pincus et al 2010). Indeed, the evidence summarised above often did not specify whether the study was investigating beliefs or behaviour, but the common measures of fear-avoidance relate to beliefs. Previous reviews also noted that many studies did not clearly differentiate between various psychological constructs, leading to terms being used interchangeably (e.g. distress and depressive mood) (Pincus et al 2002; Turk & Wilson 2010).

Limitations with the model itself have also been described in the literature. The FAM does not take into account that the course of back pain, in terms of the number of episodes experienced, pain duration, fluctuation and the impact on the individual, will vary substantially between patients (Leeuw et al 2007a). The model may only account for a particular sub-group of patients with persistent LBP rather than being applicable to all LBP populations (Leeuw et al 2007a). Most recently, Wideman et al (2013) have urged a 'rethink' of the FAM due to the proposed causal links between each of the variables not being empirically supported, and that other factors not currently included in the model, such as resilience, may potentially explain why not all patients follow the FAM's proposed pathways. Leeuw et al (2007a) also pointed out that just because

there is evidence of associations between each of the constructs in the FAM does not mean that these links are causal, and that further evidence is needed to test this.

Overall, while the FAM components may individually be predictive of LBP outcomes, the model as it currently stands may not be able to fully explain how people transition towards more persistent or recurrent trajectories of back pain. Vlaeyen et al (2009) point out that the FAM as they described it was never meant to be a finished article, and that amendments and improvements were welcomed. It is clear that, from an explanatory point of view, further work is needed to truly test the hypothesised framework of the FAM and whether the components work together in the sequence described.

1.5.2 Acceptance and Commitment Theory (ACT)

Acceptance and Commitment Therapy (ACT) is described as a type of "third wave" CBT (de Boer et al 2014), following the first wave of early behavioural approaches and second wave of cognitive-behavioural approaches. While CBT is focused on changing thoughts and behaviours (Forman et al 2012), ACT additionally focuses on living with and accepting thoughts and on the values of the patient (Gauntlett-Gilbert et al 2013). ACT has its theoretical basis in Relational Frame Theory (RFT), which assumes that cognition and language are specific learned behaviours (Hayes et al 2006). ACT has six core principles (Hayes et al 2013): *Cognitive defusion*, where people are taught to notice their negative or unhelpful thoughts rather than taking thoughts literally; *Acceptance*, which is trying to accept unwanted cognitions rather than changing them; *Being present*, which involves people focusing on themselves as they are now, not as they were in the past or could be in the future (Twohig 2012); *Values,* where the patient's goals and values are clarified as a way of keeping them motivated to change (Twohig 2012); *Self as observer*, which is "what we believe ourselves to be" (Twohig 2012, p. 503) and *Committed action,* which combines the other

processes and helps people to continue to practice them and achieve their value-linked goals (Twohig 2012). The end goal of ACT, through each of the principles, is to achieve psychological flexibility (Hayes et al 2013), which is the process of behaving in a way consistent with the persons values and goals (McCracken & Velleman 2010). Treatment involves building processes of acceptance and mindfulness using metaphors to help patients relate to therapy in a way which is more generalisable to the world outside of treatment (Hayes et al 2013). ACT is still evolving, and recently the six processes described earlier were paired up into three response styles (Vowles et al 2014): Open (defusion/acceptance); Centered (being present/self as observer); and Engaged (values/committed action), but the usefulness of these distinctions has not yet been tested.

Evidence for the key processes of ACT as mediators of outcomes is emerging, with some support for acceptance, flexibility and values (Vowles et al 2013), and avoidance and cognitive fusion (Wicksell et al 2010) being mediators of disability outcome. Acceptance has also been found to mediate outcome when not directly targeted for change (Åkerblom et al 2015), suggesting further research is needed to explore how ACT components are associated with treatment outcome.

1.5.3 Misdirected problem solving model

Eccleston & Crombez (1999; 2007) describe the misdirected problem solving model as being the result of pain demanding attention, and that in persistent pain this means that attention is constantly interrupted resulting in a vicious cycle of worry, vigilance and problem solving through repeated unsuccessful attempts to reduce pain. As was mentioned earlier, some worry and attention to pain in the short-term is adaptive (Eccleston & Crombez 2007; Flink et al 2012), but constant worry about how to stop pain may lead to hypervigilance and depression (Aldrich et al 2000; Linton & Shaw 2011). Pain is still often viewed by sufferers as a biomedical problem, leading to problem-solving behaviour around reducing the pain from a medicalised perspective (Eccleston

& Crombez 2007). If pain reduction is achieved then worry reduces, but if pain persists a "perseverance loop" is created by which the failure to reduce pain leads to further worry and further problem-solving attempts, which may again be futile. To the author's knowledge little work has been done to assess this model from a mediation perspective. One study by Flink et al (2012) examined whether framing the pain problem as a biomedical one mediated the relationship between catastrophising thoughts and biomedical-focused problem-solving behaviour in 173 patients with back, neck or shoulder pain. However, the analysis did not support this model, suggesting that the hypothesised model pathways need further examination.

1.5.4 Self-efficacy model

In the context of this thesis, self-efficacy suggests that in order to improve their outcome, patients need to take responsibility for their pain problem by trying to understand it and developing self-management strategies (Linton & Shaw 2011). People who have low self-efficacy may be less likely to develop coping mechanisms or give up more easily if there is no change in their condition (Turk & Okifuji 2002), leading to further pain and disability. Self-efficacy has been found to predict pain and disability outcomes (see Section 1.3).

Jackson et al (2014) conducted a meta-analysis to examine how self-efficacy impacts on outcomes such as impairment and pain in chronic pain populations (including, but not limited to, MSK pain) and to identify potential moderators of that relationship. Of the 83 studies they included in their review (23 of chronic LBP), older age was found to be a moderator of the self-efficacy – impairment relationship and interestingly so was the content of the measure of self-efficacy used, with studies using a measure that assessed the capacity to function despite pain having larger effect sizes than measures which evaluated pain or symptom management. The authors found that the included studies used 40 different measures of self-efficacy to correspond with the

specific context of pain they were studying. It has been suggested that the reason self-efficacy is so highly correlated with measures of pain and disability is due to the similarity between the commonly-used self-report measures of these factors (Keefe et al 2004; CJ Main, personal communication). Simply put, one would expect a high correlation between a report on what one had found difficult to do (high self-reported disability) and where one feels a lack of confidence (low self-efficacy). Several studies of observational data have shown self-efficacy to be a mediator of the relationship between pain and disability (Arnstein et al 1999; Denison et al 2004; Söderlund & Åsenlöf 2010; Costa et al 2011), and of the relationship between pain-related fear and disability (Woby et al 2007). Studies investigating self-efficacy as a mediator of the outcome of a specific treatment are scarce in the MSK pain field (see Chapter 4), although reviews exist in the fields of physical activity promotion (Baranowski et al 1998; Lubans et al 2008; Rhodes & Pfaeffli 2010), dietary behaviour change (Cerin et al 2009) and psychosocial interventions for patients with cancer (Stanton et al 2013) which all provide support for self-efficacy as a mediator of treatment outcome.

Self-efficacy is a central tenet of Social Cognitive Theory (SCT) (Bandura 1989), one of the few theoretical frameworks that actually gives information on how to influence self-efficacy using a range of techniques (Michie et al 2008). Many treatment programmes for people with persistent MSK pain focus on trying to change maladaptive behaviours that have been learned over time (Linton & Shaw 2011). Linton & Shaw (2011) also state that learning involves the individual's wider social environment, and manifested pain behaviours may partly be the result of this environment. SCT, which takes into account the context of behaviour, therefore may be important in helping to understand how these behaviours are learned and how they can be unlearned.

1.5.5 Diathesis-stress model

The diathesis-stress model suggests that individuals may be more likely to experience pain because of predisposing characteristics that interact with the personal experience of stressful events (Flor et al 1985). In other words, if a person is already under stress due to other factors, then pain might add to this pressure (Linton & Shaw 2011), or their genetics or personal history may make coping with pain more difficult for them than for others with different characteristics (Gatchel & Turk 2008). This model may help us understand the individual differences in patients with pain (Gatchel & Turk 2008). Evidence supporting this model comes from an experimental study of 51 patients who had either persistent LBP, general pain or no pain (Flor et al 1985). All patients were exposed to experimentally induced pain in four different conditions (two personally relevant stressors where the patient had to describe a pain or stressful experience, a general stressor which involved a timed mathematical task, and a neutral task which involved reciting the alphabet). Participants with persistent LBP who also experienced worry and depression experienced more muscle reactivity in the personally relevant stress conditions than those in the general or no pain groups, although the difference between the groups was small. Further evidence comes from a more recent study by Martin et al (2010) which used SEM to examine the relationships between pain and disability. They found relationships between anxiety sensitivity and fear of pain and catastrophising thoughts, fear of pain and avoidance behaviour, and avoidance behaviour and disability. The authors suggested that these results provided support for the diathesis-stress model, but their results still cannot establish the temporal order of the variables in the model (i.e. whether the variables are affected in the sequence hypothesised by the model).

1.5.6 Self-regulation model (SRM)

A final model, not described in Linton & Shaw's paper but of interest in this thesis, is the self-regulation model (SRM) (also known as the Common Sense Model). This model assumes that people develop personal ideas and perceptions about their illness, based on previous experience, which impact on how they react to it (Leventhal et al 2003). The model has five dimensions which cover how people *identify* their illness, what they see as the *cause* of it, what might happen as a result (*consequences*), how long it will last (*timeline*), and how or if they can *control* it and whether it can be cured. Our perceptions affect how we choose to cope with illness, for example whether or not we seek treatment (McAndrew et al 2008). The SRM has been described as a CBT model of physical illness rather than mental illness (Hobro et al 2004), making it potentially more relevant to pain populations where reactions are likely to be 'normal' rather than pathological. This is supported by a study which found illness perceptions to be related more strongly to physical health than mental health in a primary care population seeking treatment for a health problem (Frostholm et al 2007).

Evidence exists for illness perceptions as predictors of health outcomes (Frostholm et al 2007; Foster et al 2008; Foster et al 2010). While Frostholm et al's study examined patients with a range of health conditions, Foster et al's studies were specific to LBP. Three illness perceptions in particular (identity, control and acute-chronic timeline) were found by Foster et al to be independently predictive of disability outcome, along with self-efficacy. Patients who experienced a good outcome at six-month follow-up experienced positive changes in illness perceptions over that time period (Foster et al 2008), while those with negative changes were more likely to have a poor outcome. The authors suggested that a person's illness perceptions could explain why they respond or do not respond to a particular treatment and that changing these could potentially be important. Another study investigated the differences between people with longer- and shorter-

term LBP (van Wilgen et al 2013). They found that illness perceptions differed between the two groups, with those with persistent LBP more likely to think that the cause was physical and needed biomedical intervention and that keeping active was important.

There is also some evidence for illness perceptions being modifiable, with one randomised controlled trial (RCT) of 156 patients with persistent LBP evaluating an intervention that specifically targeted illness perceptions finding a between-group difference in illness perceptions between those who received the intervention and the waiting list control (Siemonsma et al 2011). However, this study did not find a between-group difference in their outcome of back pain-related disability, suggesting that while illness perceptions can potentially be changed if specifically targeted by an intervention, this may not necessarily lead to a change in outcome. This suggests that illness perceptions are not mediators of treatment outcome, but to my knowledge there have not been studies to formally investigate this in a MSK pain population.

1.5.7 Comparison of psychological models

While these models have been described separately, there is a degree of overlap between the constructs included in each. For example, the earliest description of the FAM (Lethem et al 1983) also talks about factors that could affect how a person reacts psychologically to pain, describing how stressful life events, previous pain experience and personal coping strategies that may have been used previously to cope with pain as key factors in whether a person avoids or confronts pain. This has some similarities with the Diathesis-stress model and Misdirected problem solving model. Also, Turk & Okifuji (2002) suggest that self-efficacy may be linked to fear-avoidance behaviour in that by having the confidence to expose themselves to a feared activity, patients may be able to reduce their fear of performing that activity. This highlights how patients may experience different 'pathways' to persistent pain or poor outcomes that may include similar

constructs. Overall, each model provides some evidence for the significance of each of the variables included in the model and the strength of their relationships with one another. However, few models have been tested in a way that allows the hypothesised pathways between the variables in the models to be investigated. In the case of the FAM, the evidence suggests that the model is in fact not working in the way it is hypothesised to with studies finding little support for the causal links between the variables. This highlights where mediation analysis could perhaps be useful in helping to examine these links.

1.6 Rationale for Thesis

To summarise, LBP globally is a costly problem in terms of time off work, healthcare utilisation, and loss of quality of life. In the UK, primary care is often the first point of call for patients experiencing LBP and this is therefore a crucial point at which patients can be identified to try and prevent further problems. Psychological factors have been found to play an important role in predicting which patients are likely to have poor outcomes and develop persistent LBP, and psychological treatments have been found to be successful in improving LBP outcomes, but it is not clear exactly how these treatments work and many only show small treatment effects. The evidence summarised above suggests that future psychological interventions for LBP need to establish what components of treatment are most effective and for which patients (Ehde et al 2014), and highlights that while several theoretical frameworks have been proposed to explain how LBP develops these models have rarely been tested using data in intervention studies. Identifying mediators of treatment effect will be important to help explain why treatments do or don't work (Kraemer et al 2002) and to identify the true mechanisms of treatment which will help to streamline psychological treatments, so that only elements that are likely to lead to positive change are included (Kazdin & Knock 2003). Day & Thorn (2014) state that there is now a

consensus that moving past efficacy to understanding treatment processes will help to maximise the value of the intervention. It is also thought that identifying mediators may help identify similarities across psychological treatments; that is, whether different treatments work through the same mechanisms (Thorn & Burns 2011; Day & Thorn 2014). Macfarlane et al (2006) state that work is needed to specifically identify which psychological factors are most important in terms of the onset and maintenance of LBP so studies can be designed to target them.

One of the reasons behind the lack of strong treatment effects of psychological treatments in previous trials may be that the most important factors for causing change in functional outcomes were not targeted during treatment. Several authors have emphasised the need for testing which psychological variables are most important, and also that different sub-groups of patients may require different factors to be targeted (e.g. Keefe et al 2004; Jellema et al 2006; Koes et al 2006; Kamper et al 2010).

Little work has been carried out in the field of MSK research related to mediation analysis, especially in terms of treatment mediation. The applied work that does exist often shows limitations in terms of design and analysis techniques, potentially leading to invalid conclusions being drawn about particular mediating factors. The importance of this thesis lies in moving forward and improving analysis and study design so that future intervention studies can include robust mediation analysis as part of a future RCT or programme of research. The next chapter will set out the overall aim and specific objectives of this thesis.

Chapter 2: Aims and Objectives

The introduction chapter defined the clinical problem and the need to improve treatment outcomes for musculoskeletal (MSK) pain (and specifically low back pain (LBP)). The concept of mediation analysis was introduced, the literature on the potential of psychological factors as treatment targets reviewed and the rationale of the thesis explained. This chapter describes the overall aim and specific objectives of the thesis, and provides an outline of the thesis.

2.1 Aim of Thesis

The overall aim of this thesis was to identify and evaluate methods and analyses used in treatment mediation analysis in MSK pain intervention studies. The relevance of this work lies in its potential to assist and improve the design, delivery and outcome of clinical trials of psychological interventions for MSK pain. The thesis will therefore begin by providing an overview of the methods of mediation analysis, explaining the different statistical approaches, assumptions often made in mediation analysis and providing recommendations regarding optimal methods for conducting mediation analysis (Chapter 3), which will be adopted where possible in the thesis.

2.2 Objectives of Thesis

Following the review of mediation analysis methods, this thesis has been split into three sections, each with its own specific objective. Each of these objectives is addressed across one or more of the thesis chapters.

2.2.1 Objective 1: What is the current evidence that psychological factors mediate disability outcome in MSK pain populations?

A systematic review was undertaken to identify previous psychological interventions in MSK populations where a mediation analysis was performed, in order to assess methods of mediation analysis and types of intervention design previously employed to investigate psychological mediators of treatment effect in MSK intervention studies (Chapter 4). Critical appraisal of study design and particularly methodology was carried out to assess whether identified mediation studies met current best practice as set out in the literature (see Chapter 3).

2.2.2 Objective 2: How can we best investigate psychological mediators of treatment effect on disability outcomes?

Observational and randomised controlled trial (RCT) data was analysed to test different elements of the framework for identifying mediators described in Chapter 3. Mediation analysis was performed across different study designs to investigate the use of different statistical approaches and identify opportunities and obstacles to using different approaches in the field of MSK pain, enabling specific recommendations for future mediation research to be made.

Observational data can be important in generating hypotheses regarding factors being potential treatment mediators, therefore observational study data was analysed to identify which of a number of psychological factors were potentially modifiable and related to change in functional

outcome in patients with back pain (Chapter 5). This did not involve conducting mediation analysis, but instead provided testable hypotheses for use with RCT data.

Data from RCTs which attempted to address psychological factors measured over at least two time points with the aim of reducing disability was then used to carry out mediation analysis. The RCTs were chosen based on their suitability for mediation analysis, as described in Chapter 3. This gave the opportunity to explore different statistical approaches to mediation analysis, including structural equation modelling (SEM) (Chapter 6) and latent growth modelling (LGM) (Chapter 7), and also allowed the testing of factors identified in Chapter 5 as potential mediators of treatment effect. Chapter 6 describes testing of some of these potential mediators using data from the STarTBack study, a large primary care-based RCT that included a psychologically informed physiotherapy intervention aimed to reduce psychological distress in patients at high risk of persistent low back pain. In Chapter 7 a more novel form of mediation analysis was carried out in another large primary care-based RCT which focused on reducing fear-avoidance beliefs in patients with low back pain (Back In Action study).

2.2.3 Objective 3: How might treatment processes be investigated using repeated assessments of key mediators and disability outcomes over time?

This final question addressed a specific aspect of study design that is seen as key to the identification of treatment mediators: the measurement of the potential mediator and outcome of interest over time. Secondary analysis was conducted on diary data obtained from secondary care patients who underwent a psychological intervention as part of their treatment for persistent MSK pain. Data on potential mediators and outcome was available both at baseline and follow-up and also during the treatment period, as patients were asked to fill in a short diary each week to help monitor their progress during treatment. This study, presented in Chapter 8,

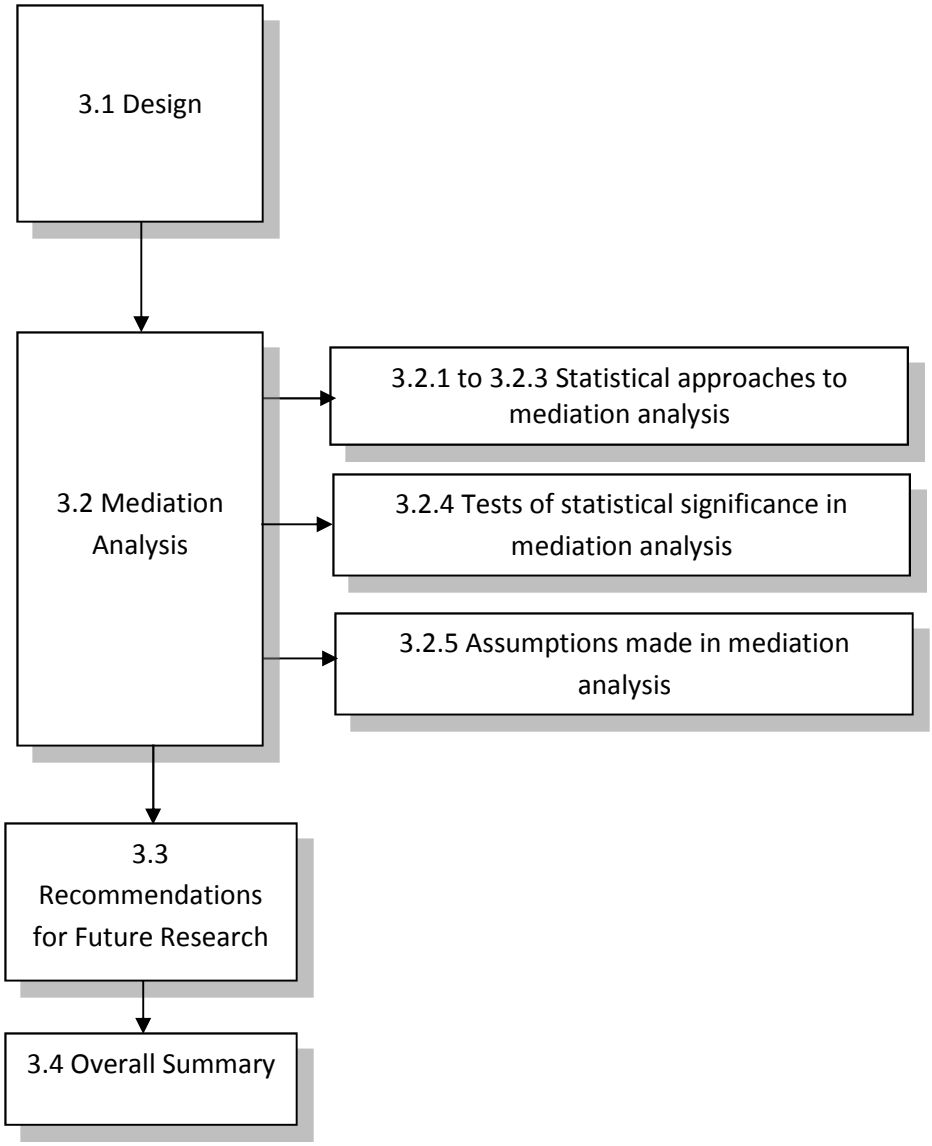
provided the opportunity to examine when particular factors of interest change and whether this provided better, more useful information than data collected only at pre- and post-treatment.

The final chapter of the thesis (Chapter 9) gives the overall discussion of the thesis, synthesising each of the empirical chapters to bring together what has been learned. It discusses the strengths and limitations of the work contained in the thesis and provides an outline of how future treatment mediation can best be conducted, highlighting the clinical and research implications of the work presented.

Chapter 3: Methodological and Design Issues of Mediation Analysis

This chapter contains an overview of the different design and methodology issues that have been outlined in relation to mediation analyses, in both observational and experimental studies. Not all of these designs confirm mediators of specific treatment effects, but each have a role in identifying potential mediating factors. Building on the definition of mediation outlined in Chapter 1, the purpose of this chapter is to describe the different design options for mediation analysis, summarise the current debate over the statistical options, and finally to derive recommendations for optimal methods of conducting mediation analysis to be employed in the thesis where possible. An overview of the sections of this chapter is given in Figure 3.1.

Figure 3.1 Overview of Chapter 3: Methodological and design issues of mediation analysis



3.1 Design

Mediation analysis has been performed using a number of different study designs. Maric et al (2012) outlined a "scientific ladder of mediation evidence" in which they proposed that while there are many types of study design that can help build evidence for mediating factors, including observational and experimental studies, some studies provide stronger evidence than others. This

section will therefore outline some of the different types of study design that have been used to carry out mediation analysis, and the problems associated with each.

Studies of mediation analysis have mostly been conducted using cross-sectional or cohort study designs. Observational studies such as these generally have the advantage of including a more representative study sample and larger sample size than experimental studies and they allow associations between variables to be established. However, confirming that a factor is a mediator of a specific treatment effect is problematic with observational data because of problems with causal inference and bias. Mediation analysis carried out on data from randomised controlled trials (RCTs) offers several benefits over observational study designs in terms of improving aspects of causality and reducing bias (Mathieu & Taylor 2006; Kazdin 2007; Nock 2007; Maric et al 2012). For example, mediation analysis in RCTs can strengthen the evidence for the mediating effect being causal because in this type of design the intervention period is clearly defined, making it easier to attribute any change in the mediator and outcome variables as being the result of a specific treatment (Mackinnon et al 2004). The inclusion of a control group as a comparison is crucial in that this gives a condition against which to compare changes in variables in the treatment group (Judd & Kenny 1981; Hill & Fritz 2011), which provides evidence for the effectiveness of the treatment above the control condition, and manipulation of a potential mediating variable within an RCT design offers a stronger basis for causal inference. However, as discussed below, there remain some problems with causality, bias and confounding.

3.1.1 Causality

In mediation analysis, it is assumed that the mediator is on a causal pathway between the independent variable (IV) e.g. treatment allocation and the dependent variable (DV) i.e. outcome. A number of authors point out the problem of establishing different aspects of causality (Maxwell

& Cole 200; Hoyt et al 2008; Gelfand et al 2009) which impact on different study designs in slightly different ways, as set out below.

An important criterion for causality is temporality, which is the timeline associated with the variables in the analysis. In order for one variable (A) to be potentially causing another (B), A must precede B in time (Rothman & Greenland 1998). In cross-sectional studies, an assessment is only provided at a single time point and therefore no start point for the first measured variable in the model can be established (Maxwell & Cole 2007), making it impossible to ascertain a causal link. A number of authors have suggested that longitudinal designs are better placed to assess change over time (e.g. Liu et al 2009; Kline 2015) and suggest that three or more waves of data may be useful for assessing temporality (Collins et al 1998; Lockhart et al 2011; Maric et al 2012; Kline 2015) despite the increase in complexity of the analysis and modelling procedures (Lockhart et al 2011). More accurate estimates of the mediating effect can be made when a number of time points are available for analysis (Maxwell & Cole 2007), but if there are not enough assessments of all potential mediator and outcome variables in cohort studies there might not be enough information to assess which variable occurred first (see Chapter 5). Hence, even in longitudinal study designs temporality cannot always be established.

It is important to consider not just the number of time points but also *when* assessments should be made. Including a large number of time points will still not help solve the issue of temporality if they miss the point at which the potential mediator and outcome change. The decisions around the timing of the assessments are very important and need to be based on the hypothesised mediation process (Collins et al 1998). Change may occur very early on or much later depending on, for example, the patient or the treatment given. In order for any mediating effects to be captured, assessments must be taken at key, clearly defined points (Laurenceau et al 2007), ideally informed by previous observational study data on the natural history and clinical course of

the condition and possibly prior research into how treatment may impact on this. An important issue highlighted here is the fact that both observational studies and RCT designs rarely include enough assessments at key time points to establish exactly when change from a specific treatment occurred (Kazdin 2007; Laurenceau et al 2007), making it difficult to determine whether change in the mediator has occurred before the change in the outcome and not vice-versa.

Plausibility (whether an association makes rational sense) is often based on what people believe to make sense rather than what is known empirically, especially if there is not much existing evidence (Rothman & Greenland 1998). Also, strong associations can be non-causal, such as the link between Down Syndrome and birth rank which was found to be confounded by the age of the mother (Rothman & Greenland 1998). A theoretical or pre-existing empirical basis for the relationship, discussed in Section 3.3.1, is therefore important in terms of justifying a proposed link.

Consistency of observations and dose-response are two important aspects of causality in the context of mediation analyses (MacKinnon 2008). Consistency suggests that repeated similar findings in different studies provide evidence of a causal link (Rothman & Greenland 1998), and dose-response is described as any effect on outcome increasing or decreasing depending on the level of the mediator (MacKinnon 2008).

3.1.2 Bias

Bias occurs as the result of error in the design of a study or measurement of the outcome variable (Sim & Wright 2000). Bias can take many forms and can also affect different study designs in different ways. The most important sources of bias in the context of mediation are selection bias and information bias.

Selection bias occurs when the probability of participants being enrolled in the study is related to the outcome of the study (Rothman & Greenland 1998). A common example is that participants are often self-selected, and there may be differences in the probability of the outcome between those who choose to take part in research and those who do not. This is true of all study designs, although in RCT designs participants are randomised once they enter the study preventing bias due to selective participation. However, there may still be participants with particular characteristics, or participants who were allocated to a particular treatment, who are more likely to drop out of the trial over time. Selection bias occurs when selective participation or drop-out is associated with outcome.

Information bias refers to errors in the collection of data that could affect the study findings (Rothman & Greenland 1998). In all study designs assessing psychological factors this can be an issue because the measurement of psychological factors needs to be via questionnaires, interviews and observations which all rely on subjective interpretation by the patient and/or observer. This precludes the provision of a 'gold standard' which measures of the factor can be compared against. Information bias is therefore hard to assess and avoid.

3.1.3 Confounding

Any association between two variables could be the result of other variables that are not measured or adjusted for in the analysis, meaning that any observed changes in the outcome cannot confidently be attributed to the variable of interest (Sim & Wright 2000). Confounding variables are those which offer an alternative explanation of a relationship between two variables (Greenland & Rothman 1998), but unlike mediating variables, confounders are not on the causal pathway under investigation. In observational designs in particular, the issue of confounding can be a significant problem.

When an RCT design is utilised, randomisation helps ensure that participants in the different treatment or exposure groups have similar characteristics at baseline and so any findings are less likely to be a result of a different distribution of confounding variables (Bullock et al 2010; Lockhart et al 2011). Randomisation therefore strengthens the argument for the treatment having a causal effect as it helps to control for confounding. A control or comparison group is also beneficial in detecting mediating effects, in that any effect of treatment given to one group, and not to other randomised (and therefore similar) groups, is more likely to be due to the treatment being given than to other factors such as the natural history of the condition (Murphy et al 2009). This can help researchers to determine whether any mediating effects would have occurred in spite of treatment or whether they are due to the treatment being given and are therefore on a causal pathway (Murphy et al 2009).

However, it has also been argued that conducting mediation analysis in the context of an RCT design does not necessarily control for all unmeasured variables that could impact on outcome (Judd & Kenny 1981), and that the problems with confounding associated with conducting mediation analysis in observational studies may therefore also apply to an extent in RCTs (Imai et al 2013). More recently this assumption has been referred to as 'sequential ignorability' (e.g. Lynch et al 2008; Imai et al 2010a; Imai et al 2010b; Ten Have & Joffe 2010) which describes how in experimental study designs it is firstly assumed that there are no pre-treatment confounders, and secondly that there are no confounders that could affect the relationship between the mediator and outcome variables. Imai et al (2010a) state that whilst the first part of this assumption does usually hold in randomised studies, the second part does not because although participants may be randomised to the intervention itself, the mediator(s) and outcome(s) are outcomes of randomisation (Lynch et al 2008; Emsley et al 2010; Lockhart et al 2011) and therefore other variables may still provide alternative explanations for any observed mediating effects. Even if an intervention sets out to directly change a mediator, the analysis cannot

determine whether any mediating effects are due to that particular factor and not to other, unmeasured variables (Bullock et al 2010). These issues can potentially be resolved by randomising at the level of the mediator variable during the intervention (MacKinnon & Dwyer 1993; MacKinnon et al 2007; Imai et al 2013). However, it is acknowledged that randomising the mediator may be a challenge, especially if there is no clear evidence for a single mediating factor (MacKinnon & Dwyer 1993) and that it can be difficult to manipulate psychological mediators (Imai et al 2013). These issues are discussed in more detail in Section 3.3.5.

While the arguments for testing mediators only in RCT designs appear strong, several key authors have opposed this conclusion. They state that randomisation and manipulation of particular mediator variables may not always be ethical (Kenny 2008) and, as noted above, RCTs are not always perfectly implemented and there may still be problems with randomisation and missing data that preclude valid conclusions being drawn (Kenny 2008; Mathieu et al 2008).

3.1.4 Summary

In terms of designing studies to best investigate mediating factors, the literature suggests that RCTs offer the highest level of evidence as they are able to better address potential problematic issues around causality and bias as discussed above. However, even RCT designs are not able to fully account for all of these issues, and more innovative designs (e.g. with potential mediators being stratified or randomised) are required to address this. As noted in the introduction to this chapter, observational and other types of intervention study design may also be useful in terms of exploring which variables are most likely to change over time and which are related to change in outcome, which can start to address issues of causality particularly when strengthened by plausible mechanisms based on previous evidence and consistent findings from preceding studies.

Section 3.2 provides an outline of mediation analysis itself. Although design and analysis are closely related, mediation analysis involves some very technical analysis challenges which are therefore discussed separately.

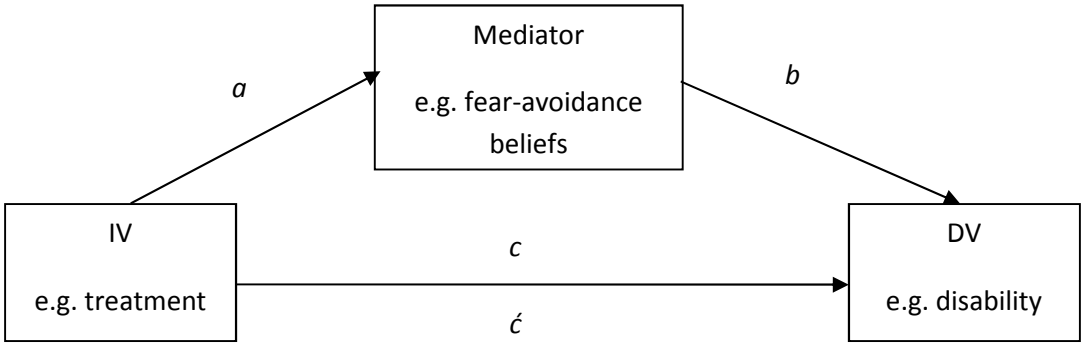
3.2 Mediation Analysis

Several different approaches to mediation analysis have been proposed, and the most appropriate method to use is currently subject to debate. A review by MacKinnon et al (2002) identified 14 different tests for mediation, which they ranked in terms of power and Type I error rates. This review discussed different approaches to mediation analysis and described several tests of the mediated effect, which are detailed below.

3.2.1 Statistical approaches to mediation analysis: Causal steps approach

A number of statistical methods for analysing mediators are available (Cheung & Lau 2008). The causal steps approach, proposed by Baron & Kenny (1986), is the most widely used method of mediation analysis. In this approach a series of regression analyses examine the separate effects of the independent variable (IV) or treatment and mediator on the dependent variable (DV) or outcome, and also the effect of the IV on the DV after controlling for the effect of the potential mediator. Three steps are proposed, outlined in Figure 3.2, which would show the presence of mediation:

Figure 3.2. Example of a simple mediation model in the Causal Steps approach



1. Demonstrate that the IV (e.g. treatment) has an effect on the outcome, DV (*c*)
2. Demonstrate that the IV (e.g. treatment) has an effect on the proposed mediator (*a*) and that the mediator has an effect on the outcome, DV (*b*)
3. Demonstrate that the effect of the IV on the DV is reduced, or becomes non-significant, after controlling for the effect of the proposed mediator on the DV (*ć*)

However, this is also one of the most heavily criticised methods of mediation analysis (MacKinnon et al 2002; Emsley et al 2010). Several authors have questioned the need for the first step to be statistically significant (e.g. MacKinnon & Fairchild 2009; Emsley et al 2010), as indirect effects can exist even if there is no overall effect of treatment on outcome (Rucker et al 2011). This means that effects could potentially be missed if this step is rigidly adhered to, and also discourages researchers from using this type of analysis to explore why the findings from a trial were negative (Emsley et al 2010; Collins et al 1998). This approach also does not impose the demonstration of a temporal order, something which was corrected in Kraemer et al's MacArthur approach (Kraemer et al 2001; 2002; 2008) which sought to more clearly define mediation.

The requirement for statistically significant relationships at each of the three steps in the Baron & Kenny approach may also reduce the power of this method to detect an effect of mediation

(MacKinnon et al 2002). Fritz & MacKinnon (2007) conducted a simulation study to test the sample size required to detect a mediating effect, and found that the causal steps approach proposed by Baron & Kenny required the largest sample size of a range of different methods tested. Others have pointed out that the third step (i.e. demonstrating that the effect of the IV on the DV is reduced after controlling for the effect of the proposed mediator on the DV) does not take into account the potentially confounding effects of unobserved variables, thus leading to inaccurate estimates (Emsley et al 2010).

Baron & Kenny (1986) distinguish between "full" or "partial" mediation, in that if the *ć* path becomes non-significant or is reduced when the mediator variable is included, this is indicative of full mediation; if the *ć* path is reduced but is still statistically significant, this suggests partial mediation. Again, a number of authors have suggested problems with these concepts. If researchers see full mediation as showing that a single variable accounts for a given relationship between two variables and that no other mediating pathways are present, this may impact on theory development (Rucker et al 2011). The expectation of full mediation may also be an unrealistic concept as this would require perfect measurement of all possible mediator variables (Rucker et al 2011). As psychological factors are by their nature latent and not directly measurable, this limitation is particularly relevant in studies involving examination of psychological factors. It is also very unlikely that any intervention would work entirely through change in a single factor for every patient.

A final problem of the Baron & Kenny approach is that it does not actually provide a test of the mediating effect; instead the effect is inferred from the results of the regression analyses (Hayes 2009) and relies on the statistical significance of the regression equations, something known to be problematic as this can be affected by sample size. The MacArthur approach also recognises this and recommends calculating an effect size (Kraemer et al 2008), which gives more information on

the magnitude of an effect in the population being examined and is therefore more clinically useful. Alternative approaches to mediation analyses outlined below (Section 3.2.2) are able to provide an estimate of the mediating effect itself.

However, there is some support for the Baron & Kenny approach. The method is said to perform as well as the other proposed approaches once the first step is removed (Hoyt et al 2008), and the analysis is easier to perform and interpret (Hoyt et al 2008). It has also been suggested that whether or not the first step in the model is included should depend on the temporal distance between the intervention and outcome variable (Frazier et al 2004). This is based on work by Shrout & Bolger (2002) who suggested that mediating effects are less likely to be detected if there is a long span of time between the intervention occurring and the outcome being measured, so the first step in Baron & Kenny's model would be unnecessary. In such cases the focus should be on the theoretical argument for a variable being a mediator (Shrout & Bolger 2002) rather than being based on statistical considerations. Others (e.g. Imai et al 2010b) have suggested that this approach can be interpreted as an average causal mediation effect so long as certain assumptions are satisfied (see Section 3.2.5). On weighing up the evidence and expert opinions overall however, it can be concluded that the Baron & Kenny approach is now somewhat outdated. Indeed, one of the original authors of the paper has also reflected that the steps outlined in the paper may have been too prescriptive, leading researchers to ignore some of the key assumptions and problems in conducting mediation analysis (Kenny 2008).

3.2.2 Statistical approaches to mediation analysis: Structural equation modelling (SEM)

Correlation and multiple regression analyses have historically been used to perform mediation analysis, with multiple regression being the method most often used to perform the causal steps approach (see Chapters 4 and 5 for summaries of the literature). The problems associated with the various steps of the causal steps approach are described above, but there are also problems

with the use of regression models in general to analyse mediators because they cannot account for measurement error, and therefore assume perfect measurement of the potential mediator variables (Brown 1997; Cheung & Lau 2008). This assumption is problematic, especially when dealing with psychological factors and self-report measures. Measurement error is a potential problem for all regression models, but especially when trying to establish whether a particular factor is responsible for the change in outcome seen, as in mediation analysis. Accurate measurement of change requires highly reliable measurement (Streiner & Norman 2008). Measurement error is magnified when looking at change, especially if a raw change score (difference between pre- and post-test) is used as there will be error in both measurements (Streiner & Norman 2008). Regression procedures also assume a causal order to the variables included in the model, which is not necessarily established (Streiner & Norman 2008).

Structural equation modelling (SEM) is a confirmatory (theory based) approach that enables the mapping out of paths or relationships between variables, taking into account error variance (Byrne 2010). SEM combines linear regression and factor analysis (Tabachnick & Fidell 2007) and models can be made up of observed and unobserved variables. The unobserved factors are latent constructs that are thought to underlie the observed (measured) variables (Holbert & Stephenson 2002). The models comprise two parts; a measurement model, examining how the observed variables relate to the unobserved (latent) factors, and a structural model, examining the relationships between the latent factors and other observed or unobserved factors. It has been recommended (Anderson & Gerbing 1988; McDonald & Ho 2002; O'Boyle & Williams 2011) that the measurement model be constructed and examined before the structural model is evaluated. This is to ensure there are no problems with the measurement model that could affect the interpretability of the structural model (O'Boyle & Williams 2011).

SEM has a number of benefits over traditional regression techniques. Firstly, measurement error can be taken into account in the model (Brown 1997; Ullman 2007). Two properties of measurement that are particularly important in mediation analysis are convergent validity (whether the items seem to represent the same construct) and discriminant validity (how distinguishable the constructs are from each other) (Mathieu & Taylor 2006). Both of these can be tested with confirmatory factor analysis (CFA) carried out within SEM, which ensures the measurement properties of the observed and latent variables in the model are examined before the structural model (examining the relationships between the latent variables) is tested. SEM also allows alternative models to be compared more easily (Mathieu & Taylor 2006). Most SEM packages also include tests of goodness-of-fit for the model, although some authors argue that the indices provided are subjective and there are no guidelines about which indices to report (McDonald & Ho 2002). This issue is discussed in more detail in Chapter 6. These benefits of SEM are evidenced in a simulation study by Iacobucci et al (2007) comparing SEM with traditional regression for mediation analysis. They found that SEM outperformed regression in terms of smaller standard errors and greater reliability of measurement with the ability to use more measures per construct.

However, SEM is not without its problems. This method generally requires a larger sample size than traditional regression analysis methods in order to obtain precise values for the specified pathways, which limits the number of studies to which this type of analysis can be applied (Quintana & Maxwell 1999). There is also no real consensus on the minimum sample size which should be used for SEM, with numbers of between 150 (Holbert & Stephenson 2002) and 200 (Tomarken & Waller 2005) being suggested. SEM also only accounts for the paths specified in the model, when in reality there could be many alternative models that are not included (Frazier et al 2004). This could lead to the goodness-of-fit indices suggesting the model is a poor fit to the data

(Quintana & Maxwell 1999). SEM also still does not test the temporal order of the variables in the model.

3.2.3 Statistical approaches to mediation analysis: Latent Growth Modelling (LGM)

In the context of mediation analysis it has been argued that the temporality of the relationship between two variables, regardless of study design, is often only an illusion as the analysis is based on between-subject rather than within-subject correlations (Collins et al 1998; Roe 2012), and that the use of two-dimensional, time-lagged models might help rectify this issue. One method that shows promise is latent growth modelling (LGM). LGM can be conducted using SEM, but differs in that the latent variables included in the LGM models represent participants' start point or baseline and their change over time, while in SEM the latent variables represent underlying constructs. LGM involves repeated measurements of observed variables to account for change over time (MacCallum & Austin 2000). LGM is becoming a more popular method of analysing longitudinal data, and could be useful in conducting mediation analysis (Lockhart et al 2011). The main strength of this approach is its ability to include variations in time (Tomarken & Waller 2005).

3.2.4 Tests of statistical significance in mediation analysis

Several methods exist to estimate the statistical significance of the indirect (mediating) effect (Mackinnon et al 2004). The most well-known of these tests is the Sobel test (Sobel 1982). This approach allows the determination of whether the potential mediator has a statistically significant effect on outcome by dividing the estimate of the effect of the mediator by its standard error and comparing this to a normal distribution (MacKinnon et al 2002). This test was recommended by Baron & Kenny (1986) to be used in conjunction with their causal steps approach as a test of statistical significance, although it has been argued that the preceding

regression analyses they propose are not necessary for this test to be performed (Hayes 2009). However, the Sobel test assumes normal distribution of the mediating effect which is frequently not the case (MacKinnon et al 2004; Hayes 2009). This is because the test is the product of two parameter estimates (*a* and *b*) (Zhao et al 2010) which may be measured on different scales and have different mean values, and also because tests of mediation are often carried out with small sample sizes (Gelfand et al 2009).

Differences in coefficients tests are used to compare the effect of treatment on outcome before and after accounting for the potential mediator (MacKinnon et al 2002; MacKinnon et al 2004). In their review, MacKinnon et al (2002) state that such tests have higher power compared to the causal steps approach but result in an increase in Type I errors compared to the other methods, meaning that a positive result may be found when in fact one is not present (Sim & Wright 2000). They also state that these tests are more useful for "simple" mediation models containing only a single potential mediator (Hayes 2009) and are more difficult to interpret in complex, multiple mediation models.

Product of coefficients approaches involve multiplying the coefficients of paths *a* and *b* together (MacKinnon et al 2004). Such tests have traditionally been more popular in sociological research, but are now frequently adopted in the psychological literature. These tests were found in MacKinnon's review (Mackinnon et al 2002) to be more conservative than the causal steps approach with lower Type I error rates, meaning that they are less likely to give a positive result if one does not exist. They are also more easily applied to models containing more than one mediator (Fritz et al 2012).

Again, these tests are not without their assumptions. It is acknowledged that the tests assume there is no measurement error in the variables being tested and that there are no omitted

variables (MacKinnon et al 2004). However, as long as these assumptions are met (see Section 3.2.5), these methods can provide accurate estimates of the mediating effect (Imai et al 2010a).

Several authors (Shrout & Bolger 2002; MacKinnon et al 2004; Gelfand et al 2009; Hayes 2009; Lockhart et al 2011) recommend that a bias-corrected bootstrap procedure should be used when assessing the magnitude of the mediating effect, as this does not assume that the effect is normally distributed. This procedure allows for nonparametric confidence intervals (CIs) to be generated, allowing estimation of the precision of the effect and of statistical significance (MacKinnon et al 2004; MacKinnon & Fairchild 2009). Bootstrapping is a process of re-sampling the data to create numerous subsamples from a single dataset (Hayes 2009; Byrne 2010). This then creates a number of parameter distributions to create a new sampling distribution (Byrne 2010). The bootstrapped estimate is the average of all of the newly created parameters (MacKinnon 2008). A bias-corrected bootstrap is particularly recommended when testing mediating effects as simulation studies have found that this method provided the lowest Type I error rate and the most power compared with other bootstrap methods (MacKinnon et al 2004; MacKinnon 2008; Williams & MacKinnon 2008). This method can be applied to complex as well as simple mediation models (Hayes 2009). The use of bootstrapping can mean it is difficult to replicate study results as the bootstrapped samples differ on each occasion, but this is not problematic if large numbers of bootstrapped samples are used (MacKinnon et al 2004; Cheung & Lau 2008).

3.2.5 Assumptions made in mediation analysis

The approaches to mediation analysis, the statistical methods available for testing mediation and the tests for the statistical significance of the mediating effect all make a number of assumptions about the data being analysed, which can be problematic if they are not adequately considered. These assumptions are also closely related to the design issues in Section 3.1. The assumptions

outlined below have been highlighted as particularly important as they may lead to biased or misleading findings.

Some mediation analysis methods, such as linear regression, do not allow for the inclusion of an estimate of measurement error in their models and therefore assume perfect measurement (Emsley et al 2010). This is not always appropriate, especially when dealing with psychological, self-report measures (Cole & Maxwell 2003; Frazier et al 2004; Kazdin 2007; Gelfand et al 2009). Other authors (e.g. Hoyle & Robinson 2004) advocate only using measures which have an internal consistency of >0.90 or using models which contain latent variables, although other authors have cautioned against spuriously high correlations as they could suggest a lack of distinctiveness between variables (e.g. Mathieu & Taylor 2006). SEM allows the inclusion of multiple indicators to provide a more accurate measure of a construct (Anderson & Gerbing 1988) and both SEM and LGM allow for the inclusion of an estimate of error in the model, although in practice few studies currently include more than one measure of a construct.

The problem of unmeasured or 'omitted' variables could also impact on the hypothesised mediation model (Judd & Kenny 1981; Frazier et al 2004; Gelfand et al 2009; Coffman 2011; Lockhart et al 2011) as these could provide alternative explanations for the mediating effect seen, or could include potential confounding factors. It is almost impossible to include measures of all possible confounding or omitted variables in a single analysis (Bullock et al 2010) and instead it has been suggested that a statement should be made that refers to this limitation in all mediation analysis studies (Gelfand et al 2009).

Any single regression or SEM/LGM analysis will only show the potential mediating pathway(s) identified by the researchers, when in fact many other models may exist. This issue is closely linked to the assumption of no unmeasured variables. The potential for alternative models to

those proposed in the study is not always acknowledged by applied researchers (Kazdin 2007; Fiedler et al 2011). SEM does have an advantage here in that models can be modified to show alternative models quite easily, although any modifications should be based on theoretical rather than statistical grounds as this will be useful in compiling evidence for a causal effect (McDonald & Ho 2002; Tomarken & Waller 2003). Alternative models may also fit the data equally as well as the hypothesised model according to the model fit statistics, so it is difficult to judge which model represents the data best (Tomarken & Waller 2003; Stone-Romero & Rosopa 2011).

A final issue with SEM/LGM analyses specifically is that while they allow multiple mediators to be included in a single analysis, they then assume independence between these different mediators which may in fact not be the case (Keele 2015). Careful consideration of each tested mediator is required to ensure they are each testing a different concept.

It is impossible to determine for certain what the direction of the relationship is when implementing correlational methods, even in the context of an experimental study. The temporal order of the variables is assumed to be correct in current proposed mediator models, when of course change in the outcome variable will still be related to change in the potential mediator whether it occurred first or second (MacKinnon 2011). The language of the approaches and methods of testing mediation however does imply causality, particularly when drawing SEM/LGM models which contain unidirectional arrows to indicate a hypothesised relationship. This has led to some applied researchers stating that their results indicate a causal relationship when in fact this assumption is not met (Quintana & Maxwell 1999).

3.2.6 Summary

This section has provided an outline of the available approaches to mediation and how they can be tested statistically. The literature suggests that the product of coefficients approach, carried

out using SEM and incorporating bias-corrected bootstrapped CIs for the indirect effect, is currently the most effective method in terms of addressing the assumptions made when conducting mediation analysis, with LGM being a promising new statistical approach to mediation analysis. Combining these recommendations with those suggested in Section 3.1 should help ensure that future mediation analysis studies provide better evidence for mediating effects. However, it is clear that even the best available methods are not without their problems and assumptions. Section 3.3 is focussed on recommendations for improvements to current best practice and outlines ways to further account for the assumptions made in mediation analysis.

3.3 Recommendations for Research

Overall, while observational studies are useful for generating hypotheses and can give information about which variables have the potential to mediate outcomes, there seems to be a move towards conducting treatment mediation analysis within the context of experimental study designs, particularly RCT designs. A number of authors from different fields have suggested ways to incorporate mediation analysis into future RCTs by making small changes to the current trial design, such as including assessments of key mediators during the intervention period. The literature concerning methods of analysis suggests that the product of coefficients approach, when conducted using SEM and including bias-corrected bootstrapped CIs, is currently considered the most robust way of performing mediation analysis.

MacKinnon (1994) and more recently Kazdin (2007) have set out guidelines concerning issues of study design and recommendations for future research. Although MacKinnon's monograph was written 13 years before Kazdin's recommendations, some of the proposed changes are very similar. Other authors (Tomarken and Waller 2003; Frazier et al 2004) have also made

recommendations as to how the mediation analysis itself should be conducted and/or reported. There is obviously overlap between some of the recommendations, and for this reason this section combines the key recommendations these authors made for improvements that apply to aspects of both design and analysis of mediation studies. My own reflections, which combine the most important recommendations given here and throughout this chapter, are given in Figure 3.3.

3.3.1 The importance of theory in mediation study design and analysis

It is important to choose mediator variables on a theoretical basis (MacKinnon 1994; Kraemer et al 2002, Frazier et al 2004; Kazdin 2007; Mathieu et al 2008) or based on previous empirical studies. Observational studies may play an important role in testing aspects of a model before particular variables are included in an RCT. The decision to conduct treatment mediation analysis should be made before the study begins, to ensure all variables to be tested as mediators can be included and measured at the right time (Kraemer et al 2002). There may be many variables in a process that help explain the effect of treatment on outcome but it is often impractical to measure all of these within a single study (MacKinnon 1994). Selecting mediators based on theory can therefore help researchers choose the key factors to focus on, which also means that clear hypotheses regarding how a treatment might work are created before the study begins (Murphy et al 2009, MacKinnon et al 2011). A theoretical basis will also help with the issue of temporality described above as this provides a framework for showing how variables might occur in a particular order (Mathieu & Taylor 2006). In addition to using an underlying theoretical framework, several authors have emphasised the importance of previous research evidence for choosing potential mediator variables to analyse (Lockhart et al 2011; MacKinnon & Fairchild 2009; Maric et al 2012). Repeating studies using different designs will add to the evidence for specific mediators and allow analysis of consistencies to add to the causal argument (Kazdin 2007).

3.3.2 Establishing temporality

Several authors (Murphy et al 2009; Kazdin 2007; Laurenceau et al 2007; Maric et al 2012; Grice et al 2015) have highlighted the importance of the timing of assessments in terms of establishing temporal precedence of the mediator before the outcome, something that is not currently routinely acknowledged in mediation research despite this being part of the definition of a mediating factor (see Chapter 1). Both the mediator and outcome should be assessed frequently over the period of time that change is expected to occur (e.g. Kazdin 2007; Laurenceau et al 2007). This again shows the importance of deciding to investigate mediation *a priori*, as having a clear hypothesis regarding when a mediator is likely to change will help researchers to plan when the optimal measurement points might be (Tomarken & Waller 2003). While studies that include assessments before and after treatment may still provide some useful information on what factors are related to outcome (Maric et al 2012), there is still a significant problem with only assessing mediator and outcome variables at pre- and post-treatment, as this does not provide evidence regarding which variable changed first (Kazdin 2007). Mediation may sometimes occur early on in treatment, meaning that mediating effects might be missed if measurements are only taken at pre- and post-treatment (Kraemer et al 2002). An intervention may also work differently for different people and the effects could occur at different times (Kazdin 2007; Laurenceau et al 2007). Multiple assessments allow exploration of this variability.

3.3.3 Psychometric properties of measures

The use of unreliable measures can increase the rate of Type II errors (the failure to detect an effect when one exists) and reduce power (MacKinnon 1994; Maric et al 2012; Frazier et al 2004; Kazdin 2007). Future studies of mediation need to take the psychometric properties of measures into account in both the design and analysis stages. A number of different psychometric properties, some of which are outlined below, are important to consider when evaluating a

measurement tool (Terwee et al 2007). All of these, where possible, should be established *a priori*.

Construct validity refers to how well a tool is measuring what it purports to measure (Terwee et al 2007). Prior to the study being undertaken, suitable validated measures should be selected. Related to this is the concept of interpretability, or how well items are understood (Terwee et al 2007). As many of the constructs to be assessed as potential mediators can only be accessed via self-report from participants it is essential that the measures make sense to those who are using them. Mathieu et al (2008) highlight this as a real issue not often considered by researchers and that construct validity should not be readily assumed.

Internal consistency relates to whether or not all the items in a tool are correlated (Terwee et al 2007). It is important that all items of a measure refer to the same underlying construct, and does not include items relating to other constructs (Streiner 2003) which would reduce its internal consistency. However, a measure with very high internal consistency may only be capturing a specific aspect of a construct, which would then impact on content validity (Streiner & Norman 2008).

Reproducibility refers to whether similar results can be obtained over time in the same participants given stability of the concept to be measured. This is important in studies of mediation where participants are likely to be followed over several time points, meaning any measurement error is magnified (Streiner & Norman 2008). Measurement error can hinder our ability to make causal inferences (Imai & Yamamoto 2010). SEM and LGM can take measurement error into account, but study designs do not always include multiple measures of variables and SEM/LGM is therefore often not used to its full potential. Any measure to be used in mediation analysis should be shown to be consistent over time.

Responsiveness is key to establishing mediating effects, as it relates to whether or not a measure can detect change over time and separate this from measurement error (Terwee et al 2007). The use of unresponsive measures in studies of mediation analysis could result in Type II errors. Floor and ceiling effects can also affect the responsiveness of a measure (Terwee et al 2007). Again, the presence of these effects should be explored in a measure prior to its inclusion in a mediation study.

3.3.4 Assessing more than one mediator

The inclusion of more than one potential mediator in studies of mediation is highly recommended (Kazdin 2007). An intervention is unlikely to work through a single mediating factor, and including several mediators makes more economical sense as it allows several tests of mediation from a single study. It is also difficult to focus on one variable in particular, especially in psychological research where the potential mediators are often subjective and closely linked to other factors (Bullock et al 2010; Campbell et al 2013b). The mediator under investigation is not necessarily also the *mechanism*, or the variable causally responsible for of the change seen (Fiedler et al 2011). Many variables may act as mediators, but only a few will actually represent causal mechanisms (Kazdin 2007; Mathieu et al 2008). However, it can also be argued that by not targeting a single mediator in particular it will not be clear which factors need to be targeted in future interventions (Bullock et al 2010). The decision of whether to include single or multiple mediators will depend on the theory being tested, the measures available and the design of the intervention.

3.3.5 Aspects of design and analysis that can facilitate mediation research

One design put forward by MacKinnon et al (2007) among others to prevent confounding as a result of unbalanced mediating variables across intervention groups in a trial involves randomising participants at the level of the mediator variable. Such a design would need to involve

randomisation of factors measured at baseline that were being manipulated, or could potentially be manipulated, during the study. This change in design is endorsed by others (e.g. Spencer et al 2005; Bullock et al 2010; Stone-Romero & Rosopa 2011; Keele 2015), who also suggest that participants should perhaps be exposed to different levels of the mediator variable to help establish mediating effects. This would be useful in establishing causality, as if a more intense condition of the mediator resulted in greater changes in the outcome, this change could be more readily attributed to the mediator than to another (perhaps unmeasured) variable. This is also known as the biologic gradient, or dose-response curve (Rothman & Greenland 1998). Two other experimental methods that may be useful in testing for mediation include blockage designs, which include a condition where the mediator is blocked in some way (MacKinnon & Fairchild 2009); and enhancement designs, which include a condition where the mediator is increased (MacKinnon & Fairchild 2009; Keele 2015). Keele (2015) gives the example of those randomised to the treatment arm of an intervention being randomly split again into two groups, where in one group the mediator is randomly encouraged. However, Keele (2015) points out that this design then requires additional assumptions, such as that the outcome is then only affected via the mediator of interest and not via other variables. Also, while these designs have been proposed as possible solutions there do not appear to be any examples of such designs actually being implemented. “Fixing” the mediator at particular levels may be impossible, especially if they relate to behaviour, meaning that while such design ideas in theory appear to be the way forward they may be difficult to apply in practice (Lynch et al 2008). It is also often difficult to manipulate psychological factors in experimental situations (Spencer et al 2005; Stone-Romero & Rosopa 2011). This issue is acknowledged as potentially the most difficult aspect of designing mediation studies (MacKinnon et al 2007).

A number of statistical methods that may strengthen the investigation of causality have been suggested, such as propensity score and instrumental variable methodology to adjust for

confounding (Coffman 2011; Rassen et al 2009; Emsley et al 2010), but these are often complex and there is some disagreement over their usefulness in the analysis of mediating factors. The assessment of mediation using the causal inference approach has recently gained attention, which involves the use of the potential outcomes framework (Imai et al 2011; Dunn et al 2013b). While these techniques are beyond the scope and focus of this PhD, future work investigating their utility is warranted. They are discussed in more detail in Chapter 9.

However, most recently even the most advanced methods of analysis have been called into question for only providing aggregate estimates, which may not be true for all individuals in the study, let alone the population to which the authors may wish to generalise the results (Grice et al 2015). Data should be observed at a very basic, individual level to pick up these differences (Grice et al 2015).

3.3.6 Back to basics?

Finally, a workshop facilitated by Mansell, Hill, Kamper and Kent at the Odense International Forum XII in October 2012 provided some insight into the opinions of MSK pain researchers regarding the future of mediation research. The general consensus at this workshop was that, aside from the points mentioned above, there was a need to return to exploratory and qualitative analyses in order to pinpoint the concepts that are most important in terms of a treatment changing outcome (i.e. the mechanisms of action). The importance of qualitative research in helping to identify key factors to test as mediators has also been advocated by other authors (Kazdin 2007), and can be especially useful in areas where no theories are available. It can be used to explore therapeutic processes and how these impact on the therapist and patient, and also why factors do or do not change and the reasons behind this (Kazdin 2007). There might be multiple outcomes of a single mediating factor, or multiple mediating factors accounting for a

single outcome, and exploratory work might be key in helping to unpick this (Kazdin 2007). Identifying key mediators is a long process, and replication of findings to give support for certain mediators being key mechanisms of outcome is an essential part of that process (MacKinnon 1994). It has been suggested that a programme of research, incorporating several different types of study designs, is needed (MacKinnon 1994; Bullock et al 2010).

3.3.7 Summary

This section has highlighted a number of recommendations from different fields on how to improve design and analysis aspects of mediation studies. There is considerable overlap between design and analysis and the recommendations can often be applied to both aspects. The key design aspects appear to involve: (1) *a priori*, theory-based decision-making around which variables could be mediating outcome; (2) assessing the proposed mediator(s) and outcome(s) frequently, during and after treatment; and (3) using well-validated measures in order to provide stronger evidence for a causal pathway. In terms of the analysis, similar recommendations are made but these also include more specific guidance including (1) the need for a power calculation for testing mediating effects and (2) testing for or acknowledging the possible presence of alternative models. However, it is important to acknowledge that many of the new ideas proposed for future research are far more complex than the design and analysis techniques currently employed and careful thought will be required in order to incorporate these new ideas into future studies.

3.4 Overall Summary

3.4.1 Design

Section 3.1 gave an overview of the designs used historically to carry out treatment mediation analysis, and outlined the problems with each type of design. Observational studies suffer from problems related to causality and bias. The general consensus from the current literature is that experimental designs in the form of RCTs provide the most robust evidence for mediating effects, and while issues with causality and bias are not totally solved, several authors suggest that making a few changes to how we conduct RCTs can support the design of studies that are more able to detect mediating effects. Suggestions for the future also indicate that the current research on mediation within the field of MSK pain may need to be brought back a few stages; several authors suggest the use of qualitative and other exploratory work to identify key factors to test as potential mediators in experimental studies.

3.4.2 Analysis

Section 3.2 highlighted the complexities in carrying out mediation analysis and the problems associated with the methods currently available. The product of coefficients approach, carried out using SEM and including bias-corrected bootstrapped CIs for the indirect (mediating) effect, is currently seen as the most appropriate method of conducting mediation analysis to examine change in potential treatment mediators. However, a number of assumptions are made by the analysis methods traditionally used which are often not met or recognised by the researchers using them, leading to potentially 'untrustworthy' mediation analysis being published in the past. The current best method described above can relax some of the assumptions and allow us to make stronger claims regarding causality, but problems with the application of SEM remain and some authors have suggested that even this method of conducting mediation analysis may still not actually provide true tests of mediating effects. More modern approaches that have been

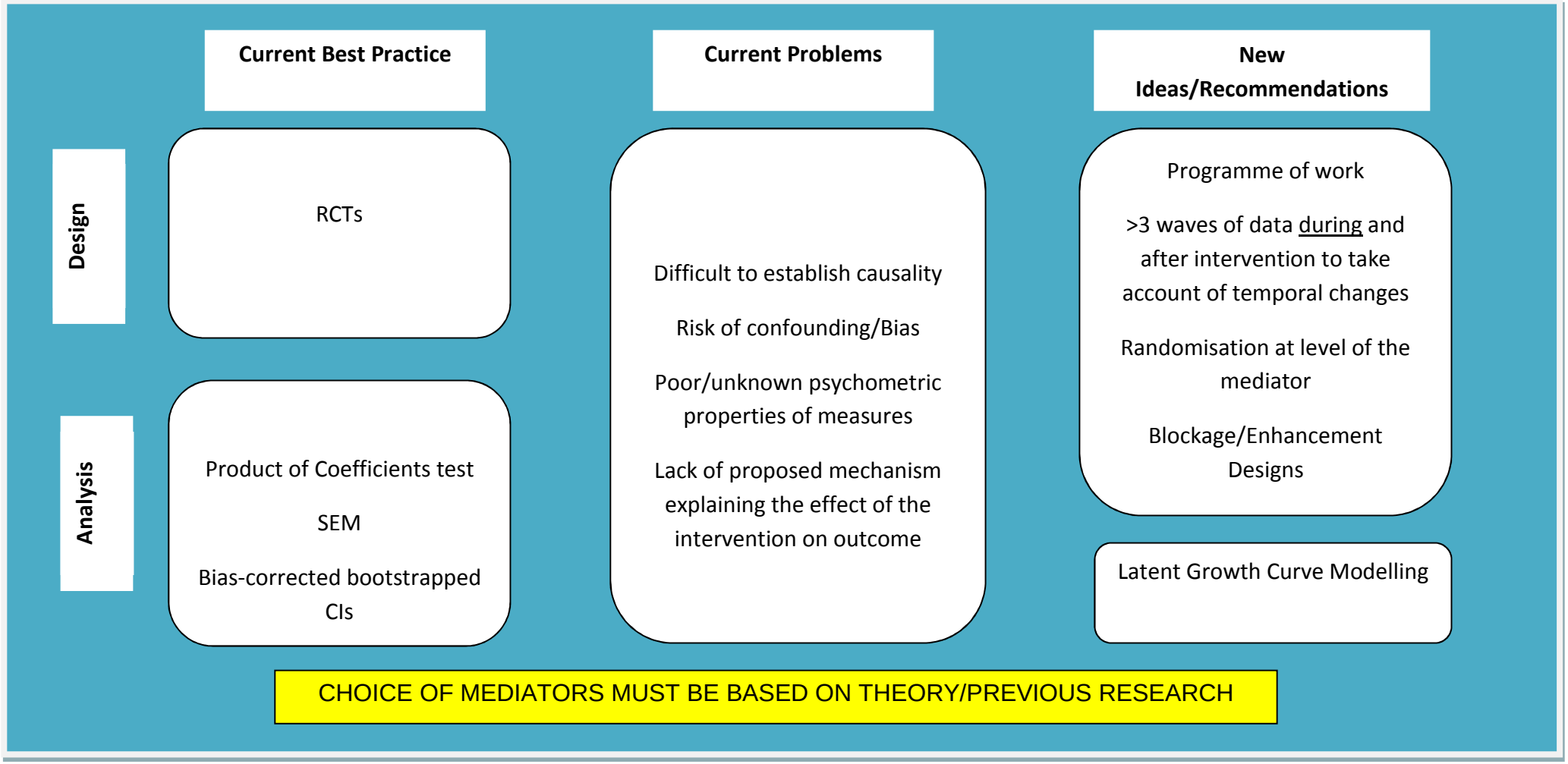
suggested recently, such as LGM, may help to solve some of the issues highlighted in this section but they substantially increase the complexity of the analysis.

3.4.3 Key recommendations

Perhaps the most important suggestion arising from this overview is that mediation analysis should be part of a programme of work that is based on sound theory to guide decision making. It is interesting to note that many of the current issues and problems with design in mediation studies, such as the timing of assessments, the inclusion of potential confounding variables and the issue of causality, can be at least partially addressed by using a strong theoretical basis. The replication of work and the use of different designs to build up evidence for particular mediating factors is also important to note, as this will help to support or refute existing theoretical models or help to build new ones. Including multiple assessments of the mediator and outcome variables during and after the intervention is also seen as key in addressing issues around causality and bias. Finally, the use of measures with sound psychometric properties or the use of analysis methods which account for measurement error are required.

This PhD thesis explored how psychological treatment mediators could be better investigated in RCTs which aimed to reduce disability in MSK pain populations. The work built on that reviewed above, focusing specifically on MSK pain, to provide further evidence for best practice in the design, conduct and analysis of studies of mediation. The current evidence for mediation studies in the MSK pain literature was firstly reviewed, followed by a series of analyses of existing datasets which tested some of the recommendations made in this chapter in order to improve and add to the current recommendations.

Figure 3.3 Current best practice, problems and recommendations regarding the design and analysis of mediators



Chapter 4: What is the current evidence that psychological factors mediate disability outcome in musculoskeletal pain populations?

4.1 Introduction

The literature review presented in Chapter 3 gave an overview of the different methods available for conducting mediation analysis and outlined the debate around the most appropriate method to use. It also highlighted the current best practice for mediation analysis and areas for improvement. An overview of the current evidence of mediation analysis in relation to musculoskeletal (MSK) pain populations is currently lacking and would be useful in terms of identifying what is already known about psychological mediators in this area, comparing different existing methodologies, and recommending areas for future research improvements. This chapter details a systematic review which has been carried out to address this gap in the research literature. An earlier version of this review was published as part of an article in Best Practice and Research Clinical Rheumatology (Mansell et al 2013).

Mediation analysis is not a new topic and has been used widely in other areas of health psychology. Systematic reviews in other fields exist, highlighting many of the issues discussed in Chapter 3. For example, Lubans et al (2008) conducted a review of mediators of physical activity in behavioural interventions for children and found seven articles that met their inclusion criteria. The numerous differences between the studies in terms of design, interventions, potential mediators, and outcomes meant that the authors struggled to come to a conclusion about specific mediators, but they did state that self-efficacy would be a good target for future interventions based on the available evidence. Another review by Cerin et al (2009) aimed to identify mediators of dietary behaviour change in adolescents and again only found seven relevant papers. The authors stated that the variety of different interventions and tested mediators made the task of

synthesising the evidence difficult, but found some evidence for self-efficacy, habit, and outcome expectations being related to outcomes. However, the authors did not feel the evidence was strong enough for these variables to be seen as treatment mediators. Finally, a recent review by Stanton et al (2013) of mediators of psychosocial interventions for patients with cancer located 16 studies which varied in terms of the type of cancer being studied and the content of the intervention. The authors identified several "classes" of mediators, including cognitive expectancies, illness representations and self-efficacy, but again highlighted several flaws with the included studies such as not basing interventions on a specific theory, using a sub-optimal method of mediation analysis and not attempting to establish the temporal relationship between the variables. These papers highlight some of the difficulties in reviewing this topic, but also the importance in bringing these problems to light so that future studies can improve the way mediation analysis is conducted.

This review identified studies of MSK pain populations where mediation analysis had been carried out on a psychological intervention (defined in Section 4.4.1). There was a need not only to provide information on constructs that have already been formally tested as treatment mediators, but also to identify which methods have been used to conduct this research. Chapter 3 identified a number of areas where mediation analysis may be flawed, potentially limiting the usefulness of previous research into mediators of treatment effect. Therefore, while the identification of variables tested previously as mediators was required, *how* each factor was identified was also very important to review.

Although the overall aim of this thesis was to identify mediators of low back pain interventions, there appeared to be a lack of studies that have conducted mediation analysis in this area specifically. The scope of this review was therefore widened to include any psychological intervention study in MSK pain populations that had conducted mediation analysis. It was thought

that this may also help identify additional treatment mediators that can be applied to the field of low back pain (LBP).

4.2 Aim

To systematically review the current evidence for psychological factors that have been tested as psychological treatment mediators in studies involving patients with MSK pain.

4.3 Objectives

1. To identify psychological constructs tested as mediators of treatment effect in intervention studies for MSK pain;

2. To assess the evidence for these constructs mediating treatment effect;

3. To identify which methods have been used to conduct mediation analysis and explore whether the choice of method may have had an impact on the study findings.

4.4 Methodology

Cochrane Review guidelines were followed to structure the review and define all its methodological processes (Higgins & Green 2011). This provides a systematic, transparent way of reporting and conducting the review so that it may be replicated.

4.4.1 Inclusion and exclusion criteria

Criteria for inclusion and exclusion were applied to each paper so that it was clear which papers should be included in the review. Details of the criteria that were applied to each paper being assessed for inclusion are given below, and summarised in Box 4.1 (page 67).

The population of interest in this review included any adult (aged 18 years or above) with non-traumatic, non-inflammatory MSK pain, such as non-specific low back pain, chronic pain or osteoarthritis. Inflammatory conditions such as rheumatoid arthritis (RA) or conditions with a known structural pathology (e.g. spinal cord injury) were excluded. Only studies that included participants with a single relevant MSK condition were included in the review, unless the paper presented analyses and results for the relevant subgroups separately. For example, if a study had included a group of patients with back pain and a group with RA, the study would be included if the groups were analysed separately so that the results for the back pain group only were given.

In this review, a psychological intervention was defined as any intervention provided by any health professional that specifically set out to change psychological factors. A control or comparator group was required so that the results from the intervention group could be compared to those of the non-intervention group.

The evidence summarised in Chapter 3 suggested that mediation analysis that provides a test of the indirect (mediating) effect conducted using structural equation modelling (SEM) or latent growth modelling (LGM) and includes bootstrapped confidence intervals (CIs) is currently seen as being able to provide the best evidence for mediation. However, as the aim of this review is to explore all the evidence of mediation analysis conducted in the field of MSK pain research, no specific type of mediation analysis was specified for inclusion. This will help address Objective 3

which involves identifying the methods used to conduct mediation analysis in MSK pain intervention studies.

In order to be included in the review, the paper must have included a measure of functional disability. This outcome is of interest because most interventions, especially those with a psychological component, aim to improve functioning rather than reduce pain, and disability is often the primary outcome of MSK trials (e.g. Burton et al 1999; Frost et al 2004). This measure could be a measure of performance, such as a timed walking test, or subjective report, using a questionnaire such as the RMDQ (Roland-Morris Disability Questionnaire; more detail about this measure is available in Chapter 5).

The recommendations for obtaining the best evidence for mediation analysis outlined in Chapter 3 suggest that intervention studies, and in particular randomised controlled trial (RCT) designs, are better able to account for the assumptions made in mediation research such as causality and temporality than cross-sectional or cohort study designs. In this review, any intervention study design was included providing that a control or comparator group was also included.

These criteria were deliberately kept broad in order to obtain the maximum number of potential studies to include. No limit was placed on the setting of the study as it was anticipated that few would have been conducted in primary care. Papers that were not available in English were excluded, as were any papers for which a full text or peer-reviewed paper was not available.

Box 4.1 Summary of inclusion criteria for studies

Adult MSK pain populations (aged 18 years or above)
The intervention should aim to change psychological factors and measure these factors over time
The study should include a comparator or control group
The study should include a measure of disability as an outcome
The study should include a method of mediation analysis and present results of the mediation analysis

4.4.2 Search methods for study identification

Four electronic databases were searched from inception to February 2012 and then again from 2012 to August 2015, using the Ovid and NHS Library interfaces (see Box 4.2). This chapter presents the combined results from the original and updated searches. Each of the databases has a slightly different focus and contains a different range of psychology and pain journals, which is why it was necessary to search more than one database when conducting the review.

Box 4.2 Electronic databases used to search literature

PsycINFO (1806-Present)
This database has a psychological focus, suggesting that it is the most likely of all the databases to pick up papers related to mediation analysis. Mediation has a number of different definitions in different disciplines so papers picked up in other databases may use terms related to "mediation" in a number of different contexts that may not always be relevant to this search. This database includes a number of journals relating to pain, as well as psychology-specific journals.

MEDLINE (1950-Present)
This is a general medical database, covering a wide variety of medical and life science journals relevant to healthcare professionals. Searching this database would identify papers published in more general medical journals rather than psychology-specific periodicals.

AMED (1985-Present)
This database covers journals relevant to complementary and allied healthcare. Again, this slightly different focus may find papers from journals not indexed in the other databases. It also covers a number of physiotherapy and spine journals which may contain relevant papers.

CINAHL (1981-Present)
This database has a focus on nursing and other allied health professions. This will also include journals that are not indexed in the previous databases that may contain papers relevant to this search.

Other databases were also available to search within the Ovid and NHS Library for Health platforms, such as EMBASE. However, after checking the journals indexed within EMBASE, it was found that many journals relevant to the search were also indexed in the databases already chosen for inclusion. EMBASE also has more of a pharmacological focus and covers a very large number of topics, suggesting that this would generate a large number of hits that may not be relevant to the search. To keep the search manageable, only the four databases described above were used.

4.4.3 Search strategy

A search strategy for electronic databases was developed based on Medical Subject Heading (MeSH) terms and free text terms for the type of study or analysis likely to have been carried out, different terms used for mediation analysis taken from previously identified key papers, and terms to identify MSK pain. The usual procedure for developing a search strategy is to include terms for the population of interest, the intervention and control groups, and the outcome.

However, while the population can be defined, no specific psychological intervention or type of control was specified. This is because there were not likely to be a large number of intervention studies that have conducted mediation analysis, so it was decided that the search should be kept as broad as possible in this area. In terms of outcome, there was also a problem with how disability was defined. When words for disability were entered as MeSH terms within each database, it became clear that this term was indexed in terms of physical or mental disability, rather than functional disability. Therefore, no specific section was added to the search to identify disability outcomes, but terms for MSK pain were entered to identify studies of patients with MSK pain which are likely to measure functional disability as an outcome, as mentioned in Section 4.4.1.

Several iterations of the search strategy were tested using key papers already identified to try and ensure that a comprehensive, yet manageable search could be carried out. The search strategy was tested in PsycINFO as this was felt to be the most relevant database for papers investigating mediation. Papers previously identified as performing mediation analysis on interventions for MSK pain patients were then searched for within the results generated by the search strategy to see if they were being picked up. Several versions of the strategy were tested in order to pick up the largest number of key papers. The final strategy for PsycINFO is shown in Box 4.3 below.

Box 4.3 Search strategy for PsycINFO

exp treatment outcome/ OR intervention$.ti,ab OR correlat$.ti,ab OR regress$.ti,ab OR "structural equation modelling".ti,ab OR "structural equation modeling".ti,ab OR "SEM".ti,ab

AND

mediat$.ti,ab OR "process variable".ti,ab OR (process adj2 evaluation).ti,ab OR (treatment$ adj2 effect$).ti,ab OR indirect effect$.ti,ab OR "indirect effect".ti,ab OR (mediat$ and analysis).ti,ab

AND

exp pain/ OR exp musculoskeletal disease/ OR musculoskeletal.ti,ab OR pain.ti,ab

LIMIT TO

Humans and English Language

4.4.4 Study selection

The results of each electronic search were downloaded into reference management software (Refworks). Duplicates were removed and the titles of all unique references were screened to remove any that were obviously irrelevant. The remaining abstracts were then obtained and any that clearly did not meet the inclusion criteria were removed. A second reviewer (Jonathan Hill) also reviewed the abstracts, and any disagreements about inclusion were discussed until consensus was achieved. If an abstract appeared to meet the inclusion criteria, or if it was not possible to tell this from the abstract alone, the full text of the article was obtained.

Some references related to conference proceedings or dissertations. Where these were found, a separate electronic search was made for publications relating to the reference. Checks for publications were conducted in Web of Knowledge using the author's name and keywords relating to the article. If no result was found, the full title of the dissertation or conference abstract was searched for electronically using Google. If no publication was found the reference was automatically excluded, as only peer reviewed publications were to be included in the review.

If a publication was found and the abstract was thought to be relevant, the full text of that publication was included as a potential paper, instead of the original dissertation or conference proceeding.

A number of papers picked up by the electronic search were protocols for intervention studies. Citation checks were therefore performed on any that met the main inclusion criteria to check whether a paper containing the analysis and results had subsequently been published.

Once relevant full texts had been identified from the literature, the reference lists of each paper were scanned for further relevant papers. Each paper was also identified in Web of Knowledge, in order to find papers that had cited the article that may also be relevant to the present review. Finally, experts identified from the literature were contacted to ask for any further papers that had not already been identified.

4.4.5 Data extraction and critical appraisal

A data extraction tool, created using Microsoft Word, was used to record the essential details from each paper. This included information on the design and setting of the study, the pain condition being examined, details of the intervention and control group, and details of the mediation analysis conducted. In studies of treatment mediation, it is important to know the details of any control or comparison used in the intervention, as this may have an effect on the potential mediator.

Critical appraisal of the mediation analysis was conducted using a tool adapted from systematic reviews of mediation analysis in behaviour change intervention studies (Cerin et al 2009; Lubans et al 2008). Many of the points in the original tool relate to aspects of mediation study design and

analysis suggested to be important in Chapter 3, but some modifications were necessary to make this tool relevant to the current topic. The original tool (Lubans et al 2008) comprised a checklist of eight items which covered areas such as whether or not the study used a theoretical framework and whether or not the psychometric properties of the measures used were tested. The authors reported that studies that scored 0-3 were low quality, studies that scored 4-6 were medium quality and those that scored 7-8 were high quality. However, this original checklist focused more on general aspects of study quality which, although important, did not really examine mediation methodology. Cerin et al (2009) adapted this tool to look more specifically at issues relating to treatment mediation, while still keeping the tool broad enough to include a range of different types of mediation analysis. They removed questions specific to Luban et al's original study question (mediators of the effect of physical activity) and replaced them with questions on whether the study tested the psychometric properties of the outcome measure as well as the mediating variable, whether steps for testing the validity of behaviour change theories had been reported and whether changes in the mediating variables preceded changes in the outcome variables. Cerin et al used the same scoring system as that used by Lubans et al.

In the present review three further questions were added to check whether a) the study tested for differences between baseline and follow-up scores in the outcome and potential mediator variables; b) whether change in the potential mediator was related to change in outcome (which were found in Chapter 3 to be important in deciding whether or not to further test a variable as a mediator); and c) whether the study controlled for potential confounding variables. This last point relates to the suggestion that even in RCTs, confounding factors can have an effect on the path between the potential mediator and outcome (e.g. Emsley et al 2010) (see Chapter 3). A summary of these changes can be found in Appendix 4.1. It was also decided that no overall score would be given for the quality of the mediation analysis, and that included studies would remain in the

review regardless of quality. This was because studies may score differently on different aspects of the critical appraisal tool. However, the scoring system was used to describe methodological strengths and weaknesses of each of the included papers. Consistency of critical appraisal which was carried out independently by two reviewers (Gemma Mansell and Jemma Cowen) was assessed by comparing scores for each of the 12 items. The modified tool used in this review is given in Table 4.1.

Table 4.1 Modified critical appraisal tool

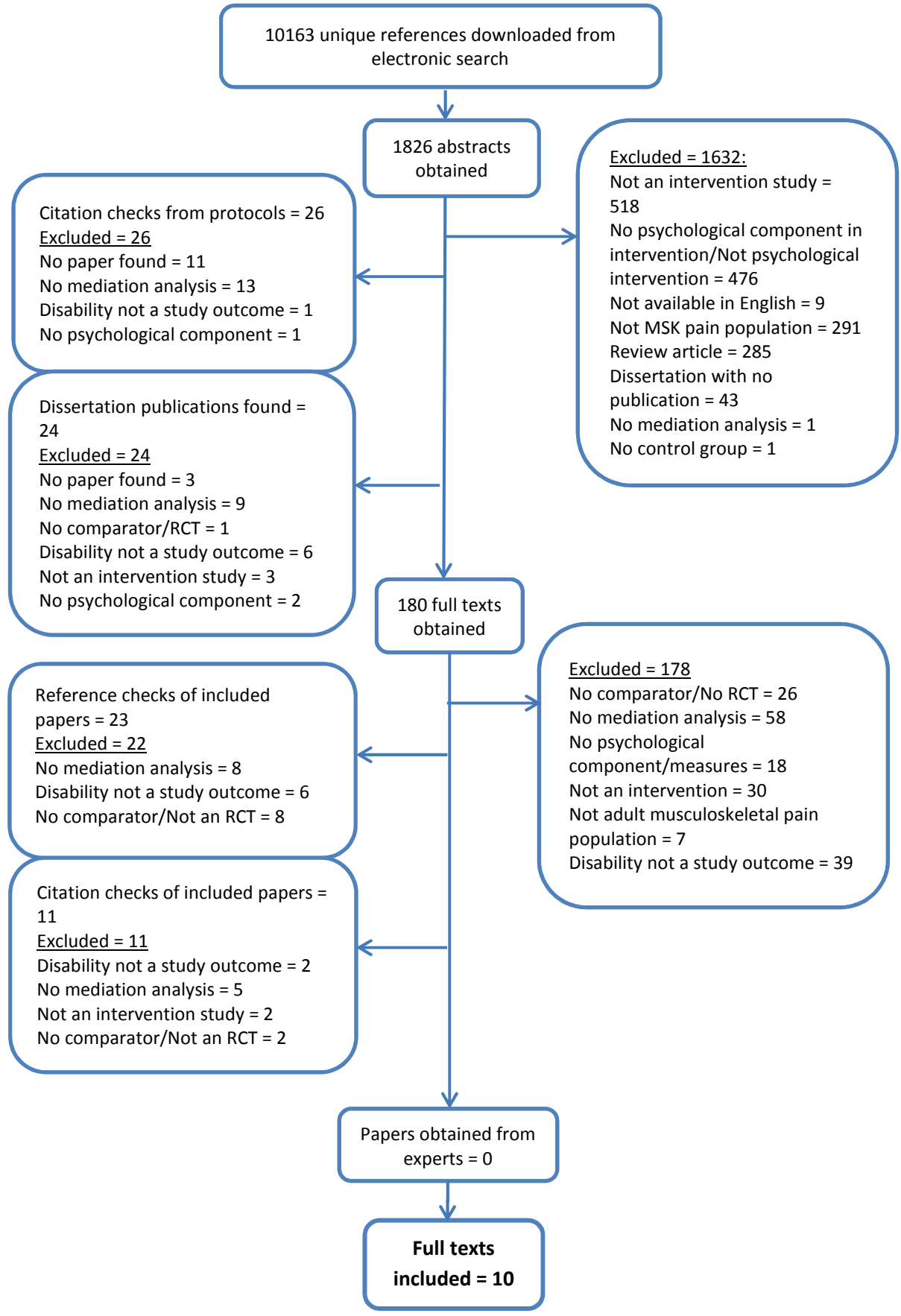
Criteria	How Item Scored	Guidance for Scoring
1. Did the study cite a theoretical framework?	Y/N	Does the study mention any psychological theory as a base for the intervention?
2. Did the study use an experimental design?	Y/N	Any design which includes an intervention and a comparator group
3. Were pilot studies conducted/ reported to test the effect of the intervention on the mediators?	Y/N	
4. Were the psychometric characteristics of the outcome measure reported AND were they within acceptable ranges (i.e. test-retest reliability or Cronbach's alpha >0.60)?	Y/N	Half a point given for each item, so that if both sub-items included the papers has a score of 1
5. Were the psychometric characteristics of the mediator measure reported AND were they within acceptable ranges (i.e. test-retest reliability or Cronbach's alpha >0.60)?	Y/N	Half a point given for each item, so that if both sub-items included the papers has a score of 1
6. Did the study report a statistically significant change between baseline and follow-up for the potential mediator?	Y/N	
7. Was the change in the potential mediator correlated with change in outcome?	Y/N	
8. Were the study methods/procedures designed to influence mediating factors?	Y/N	A statement or hypothesis that the intervention procedure focuses on affecting the potential mediator(s)
9. Did the study report a power calculation, and was the study powered to detect mediation?	Y/N	A statement of a power calculation reported in the paper
10. Were statistically appropriate/acceptable methods of data analysis used?	Y/N	A statement that the test of the indirect effect was performed using the product of coefficients approach and included bootstrapped CIs
11. Did the study control for possible confounding factors?	Y/N	
12. Did the study ascertain whether changes in the mediating variable preceded changes in the outcome variable?	Y/N	

The data extraction and critical appraisal tool were combined into a single instrument (see Appendix 4.2 for the template). This was piloted on several papers external to the review to ensure all necessary information was being recorded and the two reviewers agreed on the interpretation and judgment of items.

4.5 Results

When the electronic search was conducted, 10,163 unique references were downloaded. The titles of these references were scanned, and 1826 abstracts of articles thought to be possibly relevant were obtained. A number of these abstracts (50) were found to be either protocols or thesis dissertations, none of which were found to be relevant. After reading through all of the abstracts, 180 full papers were retrieved and nine papers were found to meet the inclusion criteria. Reference and citation checks were carried out on these papers and one further paper was identified from the reference check, resulting in a total of 10 papers being included. No further papers were identified through contact with experts. A flowchart outlining how studies were excluded at each stage of the review process is given in Figure 4.1. The studies were heterogeneous in nature with different study populations being recruited, different methods of mediation analysis used and a range of different interventions being investigated. In all but five studies (Nicassio et al 1997; Smeets et al 2006; Leeuw et al 2008; Robinson et al 2013; Wicksell et al 2013) the mediation analysis was carried out as secondary analysis or as a secondary objective, rather than being the primary focus of the study.

Figure 4.1 Flowchart presenting search results and selection of studies



4.5.1 Summary of studies

The majority of the studies (Nicassio et al 1997; Focht et al 2005; Turner et al 2007; Seymour et al 2009; Robinson et al 2013) were conducted in the USA. Three (Spinhoven et al 2004; Smeets et al 2006; Leeuw et al 2008) were conducted in the Netherlands and two (Wicksell et al 2010; Wicksell et al 2013) were carried out in Sweden.

Four studies (Smeets et al 2006; Turner et al 2007; Leeuw et al 2008; Wicksell et al 2013) were conducted in secondary care, and the remaining papers examined community populations or patient groups (for example, Wicksell et al 2010 recruited patients from an organisation for patients with Whiplash Associated Disorder (WAD)).

The papers also focused on a range of painful MSK conditions. Three (Spinhoven et al 2004; Smeets et al 2006; Leeuw et al 2008) investigated patients with chronic LBP. The other papers examined fibromyalgia (Nicassio et al 1997; Wicksell et al 2013), WAD (Wicksell et al 2010; Robinson et al 2013), Temporomandibular Disorder (TMD) (Turner et al 2007), knee osteoarthritis (OA) (Focht et al 2005) and hip and/or knee OA (Seymour et al 2009).

The number of participants involved in the studies ranged from 21 in Wicksell et al's (2010) study to 351 in Seymour et al's (2009) study. Several studies included more than one active intervention group (Spinhoven et al 2004; Focht et al 2005; Smeets et al 2006; Leeuw et al 2008). Interventions lasted for between three and 12 weeks with the exception of Focht et al (2005), in which the intervention lasted for 18 months. Length of follow-up was reported as four months (Wicksell et al 2010; Wicksell et al 2013), six months (Nicassio et al 1997; Seymour et al 2009) or 12 months (Spinhoven et al 2004; Turner et al 2007; Leeuw et al 2008), and the remaining three studies did

not follow patients up after the intervention had ended. Table 4.2 gives a summary of the included studies.

Table 4.2 Summary of papers included in the systematic review

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Focht et al 2005	USA; Community; RCT Patients with radiographic evidence of knee OA aged 60 years and over; BMI of 28 or over; self-reported knee pain on most days; less than 20 minutes exercise a week in the last 6 months; self-reported difficulties with daily activities	Exercise (*n*=80): Aerobics and resistance training Dietary weight loss (*n*=82): Based on group dynamics and social cognitive theory – education, self-regulatory skills training Combination therapy (*n*=76): Completion of both programmes	Healthy Lifestyle (*n*=78): Designed to provide an attention control	6 months, 18 months	Mediators: Walking self-efficacy, Stair-climbing self-efficacy Outcomes: Mobility (walking and stair climbing), Pain	ANCOVA	Stair-climbing self-efficacy was reported to mediate performance-based mobility in the Combination therapy group (ANCOVA result $F_{(3,192)}$=2.78, *p*=0.0425); Unclear which time point was used in the change scores

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Nicassio et al 1997	USA; Community; Randomised trial Patients with diagnosed fibromyalgia (63f, 8m); average age 53.1 (range 24-78)	Behavioural treatment (*n*=48): Education, relaxation training, goal-setting, support person	Education (*n*=38): Lectures, group discussion	Post-treatment (10 weeks); 6 months	Mediators: Helplessness, Pain coping, Social support Outcomes: Disability, Depression, Pain (observed and self-reported), Myalgia	"series of simple regression analyses"	Helplessness and passive coping were not found to mediate disability in either the treatment or control group (no statistics reported for this outcome)

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Leeuw et al 2008	Netherlands; Secondary care; RCT Patients with chronic LBP; mean age 45.32; 51.8% male	GEXP (Graded Exposure in vivo) (*n*=42): Establishment of a graded hierarchy of fear-inducing activities and gradual exposure to these activities starting with the least fearful	GA (Graded Activity) (*n*=43): Aimed to improve physical function by positively reinforcing healthy behaviour and activity goals set by the patient	12 months	Mediators: Pain catastrophising, perceived harmfulness of activities Outcomes: Functional disability, main complaints	Joint significance test	"The main effect of EXP relative to GA on functional disability was significantly mediated by decreases in pain catastrophising and perceived harmfulness of activities, as demonstrated by a marginal effect of treatment on outcome once both mediators were included in the model"

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Robinson et al 2013	USA; Community physician/self-referral; RCT Patients with WAD grades I and II	Exposure therapy (*n*=70): Patients asked to rate fear-provoking activities from least to most fearful, then over the treatment period asked to first imagine then actually perform those activities, starting with the least fearful	Information booklet only (*n*=57): Booklet given on whiplash injuries and associated pain – information given on gradual increase of activity but no emphasis on exposure Information booklet plus didactic lecture (*n*=64): In addition to booklet, patients received educational sessions from a psychologist, physiotherapist and physician	Post-treatment (within 10 days of final treatment session)	Mediator: Fear of specific neck movements Outcome: Neck disability	Product of coefficients, 95% bootstrapped CIs	Fear of activity was found to significantly mediate the effect of treatment on outcome (-3.82, 95% CI 1.70 to 6.34)

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Seymour et al 2009	USA; Community; Before and after study Patients with hip/Knee OA; no complicated surgery or steroid injections in previous 6 months; mean age 71; 80%f	Fit & Strong! Programme (exercise therapy) delivered by physiotherapists	Same programme, delivered by certified exercise instructors	Two months, six months	Mediators: Self-management elf-efficacy (exercise, pain management, symptom management), Exercise adherence self-efficacy (barriers adherence, time adherence), Attendance, Exercise maintenance Outcomes: Function (muscle strength and exercise capacity, physical function), Pain	Random effects model	Exercise adherence and self-efficacy mediated disability outcomes in physiotherapy delivered care at two months (main effect 0.246, 1.49 (0.135); interaction 1.12, 3.44 (0.001) but not at six months

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Smeets et al 2006	Netherlands; Primary care; RCT Participants with chronic non-specific LBP (three months or more in duration); aged between 18-65	Cognitive-behavioural therapy (*n*=55): Problem-solving training and graded activity Active Physical Therapy (*n*=52): Aerobic and endurance exercise Combined Therapy (*n*=55): Combination of both therapies	Waiting List (*n*=49)	Post-treatment (10 weeks)	Mediators: Pain catastrophising, Internal pain control Outcomes: Disability, Pain, Depression, Patient-specific complaints	Baron & Kenny, plus Sobel test	Internal pain control not found to mediate disability in any treatment condition; Pain catastrophising found to mediate disability in all treatment conditions (Sobel test = -2.754, *p*=0.006 for APT; -2.269, *p*=0.023 for CBT; -2.278, *p*=0.022 for CT)

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Spinhoven et al 2004	Netherlands; Primary care; RCT Patients with chronic LBP (six months or more duration); 54m and 94f; mean age 39.8 (range 18-64)	Operant Behavioural Treatment with Cognitive Coping Skills (*n*=59): Enhancing self-control and relaxation skills, plus homework for reinforcement Operant Behavioural Treatment with Group Discussion (*n*=58*):* Aimed to provide an attention-control for other intervention	Waiting List (*n*= 31)	12 months	Mediators: Catastrophising, Pain coping, Internal pain control, External pain control Outcomes: Activity tolerance, Pain (intensity and behaviour), Depression	Baron & Kenny	Catastrophising found to partially mediate activity tolerance (magnitude of semi partial correlation (-.15 significantly (p<0.05) smaller than zero-order correlation (-0.40). Unclear how this relates to mediation. Treatment conditions were collapsed into one group for analysis so not clear which intervention resulted in the mediating effects.

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Turner et al 2007	USA; Secondary care; RCT Patients with diagnosed TMD *(no further detail)*	Cognitive-Behavioural Therapy (*n*=55): No detail given	Education/ Attention (*n*=60): No detail given	Post-treatment (eight weeks), six months, 12 months	Mediators: Self-efficacy, Pain beliefs (disability, harm and control), Catastrophising, Coping (Relaxation) Outcomes: Jaw use limitation, Pain (intensity and activity interference)	Baron & Kenny, plus Sobel and bootstrapping	Pain beliefs, coping and catastrophising mediated activity interference at 12 months in the CBT group (disability subscale (indirect effect = -1.18, CI -1.80 to -0.78), harm subscale (indirect effect = -0.70, CI -1.22 to -0.34) and control subscale (indirect effect = -1.18, CI -1.87 to -0.56)), pain catastrophising (indirect effect = -0.59, CI -1.11 to -0.31),coping (rumination subscale (indirect effect = -0.23, CI -0.61 to -0.01) and self-efficacy (indirect effect -0.87, CI -1.55 to -0.49).

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Wicksell et al 2010	Sweden; Community – WAD support group; RCT Patients diagnosed with whiplash-associated disorder (WAD) of three months or more duration; aged 20 years or over	Acceptance and Commitment Therapy (*n*=11): Acceptance as alternative to avoidance, assessment of individual values, graded activity	Treatment as usual (*n*=10)	Post-treatment (eight weeks), four months	Mediators: Pain intensity, Anxiety, Depression, Self-efficacy, Fear-avoidance, Psychological inflexibility Outcomes: Pain disability, Life satisfaction	Baron & Kenny, plus cross-product of coefficients and bootstrapping	Psychological flexibility mediated disability at post-treatment but not at four months follow-up in the ACT group (bootstrapped indirect effect = 7.69, CI 0.24-17.19)

Reference	Country, Setting & Design, Study Population	Intervention(s)	Control	Follow-up	Mediator(s) and Outcome(s) Tested	Mediation Analysis Method	Conclusion
Wicksell et al 2013	USA; Secondary care, Fibromyalgia All female patients; mean age 45.1 years	Acceptance and Commitment Therapy (*n*= 23): 12 weekly group sessions on preparing for behavioural change, shifting perspective to increase functioning despite pain, and identifying behavioural goals	Waiting list	Four months post-treatment	Mediator: Psychological inflexibility Outcome: Pain disability	Product of coefficients with bootstrapped 95% CIs	Psychological inflexibility was found to mediate the relationship between treatment and disability (-3.65, 95% CI -11.52 to -0.54)

4.5.2 Summary of critical appraisal

Two reviewers (Gemma Mansell and Jemma Cowen) used the critical appraisal tool to assess each study. Overall, complete agreement was achieved with three of the included studies; in the remaining papers disagreements occurred with either one or two of the 12 critical appraisal items (items 1, 4 or 7 in Table 4.1). These discrepancies were discussed between the reviewers and the papers were re-examined in order to achieve consensus. The quality of the studies in terms of mediation analysis was poor overall. The key points for design and analysis, and the overall critical appraisal of each of the studies, are given below.

Cognitive-behavioural therapy (CBT) was cited most often as a basis for the intervention (Spinhoven et al 2004; Smeets et al 2006; Turner et al 2007; Leeuw et al 2008; Wicksell et al 2010; Robinson et al 2013; Wicksell et al 2013). Basing the intervention on a theoretical framework (item 1) is seen as a key area for improvement in mediation analysis studies (e.g. Kazdin 2007) (see Chapter 3). However, while CBT is based on principles of cognitive-behavioural theory, the techniques used to treat patients vary widely. For example Spinhoven et al (2004), Smeets et al (2006) and Leeuw et al (2008) used operant behavioural techniques in their trials, but while the Smeets and Leeuw trials focused on aspects of graded activity, the Spinhoven trial only mentioned trying to increase "healthy" behaviours without giving specific detail as to how this was done. Turner et al's (2007) trial focused on pain management training. Wicksell et al's trials were based on a specific form of CBT (Acceptance and Commitment Therapy (ACT)). They describe ACT as an "exposure-based behaviour therapy" (p.2) where patients are taught to accept negative thoughts without having to alter personal values. ACT also does not aim to reduce patients' levels of pain or distress (Wicksell et al 2010). Focht et al (2005) cited social cognitive theory as the basis for their intervention with a focus on increasing self-regulatory skills, but did not explain the model or which aspects were key to the intervention. The final papers (Nicassio et al 1997; Seymour et al

2009) did not specifically state a particular intervention, but Nicassio et al did cite previous studies based on cognitive-behavioural principles and their intervention did include giving information based on the gate control theory of pain. Seymour et al's study was the only paper not to mention any type of theoretical basis despite including self-efficacy as a potential mediator, which is a key component of social cognitive theory. The variety of techniques and principles employed in each of these studies and the different variables that were chosen as potential mediators to investigate highlights the potential difficulties in using broad theoretical frameworks and not deciding *a priori* how the theory might work in an intervention context.

Smeets et al (2006), Leeuw et al (2008) and Wicksell et al (2013) reported information on the psychometric characteristics of both the potential mediator measures (item 4) and outcome measures (item 5). Focht et al (2005) and Wicksell et al (2010) reported data regarding psychometric properties for their outcome measures only and Spinhoven et al (2004) and Robinson et al (2013) only reported psychometric properties of their mediator measures. Only Smeets et al (2006) and Spinhoven et al (2004) reported tests of internal consistency within their study population, rather than using values taken from other studies. These were all found to be above 0.60 for Cronbach's alpha as specified in the critical appraisal tool, indicating good internal consistency.

In terms of the mediation methodology used, all the included studies were designed to test the influence of potential mediator variables on treatment outcome, and all reported that patient scores on the mediator measures changed between baseline and post-treatment/follow-up but only four (Nicassio et al 1997; Spinhoven et al 2004; Wicksell et al 2010; Wicksell et al 2013) reported correlation analysis to show that this change was associated with change in the outcome measure (item 7). This suggests that in some studies, the variable's potential to mediate outcome may not have been adequately assessed prior to the mediation analysis being conducted.

Only four papers conducted mediation analysis that has been shown to be appropriate according to current evidence (Wicksell et al 2010; Wicksell et al 2013; Turner et al 2007; Robinson et al 2013) in that tests of indirect effect were conducted with non-parametric bootstrapping, as the indirect effect is rarely normally distributed (item 10). However, only Turner et al (2007) reported that their analysis was carried out using SEM. No study reported a power calculation to show that the study was powered to detect mediating effects (item 9) and many studies recruited small samples (for example, only 21 participants were included overall in Wicksell et al's 2010 study). This suggests that the samples may not have been large enough to adequately perform mediation analysis.

Acknowledging the influence of other measured variables is also seen as important in establishing the effects of the variables being tested as mediators (see Chapter 3). Only four studies (Smeets et al 2006; Focht et al 2005; Leeuw et al 2008; Robinson et al 2013) reported that they controlled for potential confounding variables in their mediation analysis (item 11). A summary of the results of the critical appraisal can be found in Table 4.3.

Table 4.3 Summary of the results of the critical appraisal of mediation analysis

Reference	1 Intervention based on theory	2 Experimental design used	3 Pilot study conducted	4 Psychometric characteristics of outcome measures reported	5 Psychometric characteristics of mediator measures reported	6 Change recorded between baseline and follow-up in potential mediators	7 Change in mediator correlated with change in outcome	8 Study designed to influence mediating variables	9 Study powered to detect mediation	10 Appropriate mediation analysis conducted	11 Confounders controlled for	12 Ascertainment of change in mediator(s) before change in outcome
Focht et al 2005	Y	Y	N	Y	N	Y	N	Y	N	N	Y – Baseline values	N
Leeuw et al 2008	Y	Y	N	Y	Y	Y	N	Y	N	N	Y - Treatment	N
Nicassio et al 1997	N	Y	N	N	N	Y	Y	Y	N	N	N	N
Robinson et al 2013	Y	Y	N	N	Y	Y	N	Y	N	Y	Y – Age and attorney involvement	N
Seymour et al 2009	N	Y	N	N	N	Y	N	Y	N	N	N	N
Smeets et al 2006	Y	Y	N	Y	Y	Y	N	Y	N	N	Y – Age, gender, treatment centre, baseline scores of mediator and outcome variables	N
Spinhoven et al 2004	Y	Y	N	N	Y	Y	Y	Y	N	N	N	N

Reference	1 Intervention based on theory	2 Experimental design used	3 Pilot study conducted	4 Psychometric characteristics of outcome measures reported	5 Psychometric characteristics of mediator measures reported	6 Change recorded between baseline and follow-up in potential mediators	7 Change in mediator correlated with change in outcome	8 Study designed to influence mediating variables	9 Study powered to detect mediation	10 Appropriate mediation analysis conducted	11 Confounders controlled for	12 Ascertainment of change in mediator(s) before change in outcome
Turner et al 2007	N	Y	N	N	N	Y	N	Y	N	Y	N	N
Wicksell et al 2010	Y	Y	N	Y	N	Y	Y	Y	N	Y	N	N
Wicksell et al 2013	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	N

Y = Yes; N = No

4.5.3 Results of the mediation analysis performed in each study

The majority of the included papers stated that they had used the Baron & Kenny (1986) causal steps approach to conduct their mediation analysis, or had used some other form of regression analysis. Smeets et al (2006) reported using Baron & Kenny's method to conduct their mediation analysis, and included the Sobel (1982) test to examine the indirect effect. This study tested whether catastrophising thoughts and internal pain control were treatment mediators of functional disability (among other outcome measures) in 211 patients with low back pain who took part in either cognitive-behavioural therapy (CBT), active physical therapy (APT), a combination of CBT and APT, or a waiting list control group. Patients in the APT group took part in aerobic and endurance exercises, while those in the CBT group received problem-solving training and graded activity exercises. All treatments lasted for 10 weeks. Measures of the outcome and mediator variables were taken at baseline and post-treatment (as soon as the 10-week treatment period had ended). The authors found that internal pain control did not change significantly over the 10-week period in either the CBT or APT treatment, and therefore did not test this variable further. Experts (e.g. MacKinnon et al 2007) do state that if a variable does not change over time (i.e. is not modifiable) then it is unlikely to be a potential mediator, so this decision by Smeets et al is supported by expert opinion. Pain catastrophising was found to be a mediator of change in disability, with significant ($p<0.05$) Sobel *t*-test values reported in all three active treatments. The authors suggested that this finding could be due to the APT helping to show participants that they were able to participate in activities that they previously felt they could not do, which led to a reduction in their catastrophising scores. They also suggested that the changes in catastrophising seen in previously reported studies of CBT interventions may actually be due to other aspects of treatment.

A similar study was conducted by Leeuw et al (2008), which examined mediators of the relationship between two CBT-based treatments (exposure in vivo and graded activity) and disability outcome in 85 patients with chronic LBP. Exposure in vivo involved gradual exposure to a set of fearful activities, graded by the patient from least to most fearful. Graded activity, based on operant behavioural techniques, aimed to improve function in patients through positive reinforcement of activity goals set by the patient. Pain catastrophising and perceived harmfulness of activities were assessed as mediators of both treatments, with mediation analysis being conducted via the joint significance test, which Leeuw et al describe as the significance of the association between the treatment and mediator (*a* path) and mediator and outcome (*b* path) adjusting for treatment. This study found that exposure in vivo was not significantly more effective than graded activity in affecting disability outcome, but that pain catastrophising and perceived harmfulness were mediators of the difference between the two treatments (graded exposure had more of an effect on these two mediators than graded activity). This study is strengthened by its adjustment for baseline values of the mediating variables, but the reliance of the mediation analysis on statistical significance is problematic and no actual mediating effect is provided.

Wicksell et al (2010) also used the Baron & Kenny method, but included the cross-product of coefficients approach and obtained bootstrapped confidence intervals (CIs) to test for the mediating effect. Bootstrapping, the process of resampling the original data many times over, helps to account for the often non-normal distribution of the mediating effect estimate (see Chapter 3 for a more detailed explanation). This study compared Acceptance and Commitment Therapy (ACT) plus treatment as usual to only receiving treatment as usual in a group of 21 patients with WAD recruited from a community WAD support group, and tested a number of potential mediators; pain intensity, anxiety, depression, fear-avoidance beliefs, self-efficacy and

psychological inflexibility. Change scores were calculated for the outcome measures but in the analyses only the post-treatment (eight-week) scores of each mediator variable were used. The authors carried out analyses to look at mediation between pre-treatment and post-treatment change in the outcome measures, and between pre-treatment and four-month follow-up change in the outcome measures. They found that psychological inflexibility was the only variable to significantly mediate change in disability, but the confidence intervals of the effect were very wide (7.69, bootstrapped 95% CI 0.24 to 17.19), and this only occurred between pre-treatment and post-treatment; the effect was not present by four-month follow-up. Also, the authors stated that they used post-treatment scores of the potential mediators in their analysis rather than change scores because they expected psychological flexibility to predict positive outcomes. The more recent included paper by Wicksell et al (2013) reported conducting analysis in the same way as that carried out by Wicksell et al (2010), but this later study included only female patients with fibromyalgia and tested only psychological inflexibility (pre- to post-treatment change) as a mediator of disability outcome (pre-treatment to four-month follow-up change). ACT was compared with waiting list control group, with 12 weekly group sessions of ACT being prescribed. They found that change in psychological inflexibility did have a statistically significant mediating effect on the relationship between treatment allocation and change in disability outcome (-3.65, bootstrapped 95% CI -11.52 to -0.54), but the wide CI again suggests a lack of precision.

Robinson et al (2013) also used the product of coefficients approach with bootstrapped 95% CIs to test reduction in fear of activity as a mediator of the relationship between treatment allocation (exposure therapy versus an information booklet versus information booklet plus didactic discussion) and disability outcome in 191 patients with WAD. The exposure intervention was carried out weekly for three weeks. Patients were asked to first imagine, then actually perform a set of activities they had previously graded from least to most fearful over the three sessions. Fear

of activity was found to significantly mediate the relationship between exposure therapy and disability outcome (3.82, bootstrapped 95% CI 1.70 to 6.34).

Turner et al (2007) also obtained bootstrapped 95% CIs for their estimates of the mediating effect. This was the only included study to report the use of structural equation modelling (SEM), although no latent variables were used in the analysis. This study compared CBT to an education control group over an eight-week period for the treatment of temporomandibular disorder (TMD) (*n*=115), and investigated a range of mediators. They found that self-efficacy (-0.87, bootstrapped 95% CI -1.55 to -0.49), pain beliefs (disability subscale (-1.18, bootstrapped 95% CI -1.80 to -0.78), harm subscale (-0.70, bootstrapped 95% CI -1.22 to -0.34) and control subscale (-1.18, bootstrapped 95% CI -1.87 to -0.56)), pain catastrophising (-0.59, bootstrapped 95% CI -1.11 to -0.31) and coping (rumination subscale (-0.23, bootstrapped 95% CI -0.61 to -0.01)) were mediators of jaw use limitation. However, despite being fairly precise estimates according to the reported confidence intervals these only appear to be small coefficients, although it should be noted that there is no clear guidance on how big an effect should be before it is judged as 'large' in the context of mediation analysis (Kelley & Preacher 2012). No details were given in the paper about the content of either intervention, although a previous paper (Turner et al 2006) stated that the CBT involved pain management training, delivered by a psychologist, to improve patient-specific activity goals. The education group involved a self-care management programme and was designed to act as an attention control for the CBT condition.

Spinhoven et al (2004) also used the Baron & Kenny method to conduct their analysis in a sample of 148 patients with chronic LBP. In this study all patients received operant-behavioural therapy (OBT) over 10 weeks, but 59 were randomised to receive cognitive coping skills treatment in addition to OBT, 58 were randomised to group discussion in addition to OBT and the remaining 31

were randomly allocated to a waiting list control group. The coping skills training involved patients undergoing three separate phases; one to enhance their self-control, one to teach relaxation techniques and finally reinforcement of these skills via homework assignments. The authors hypothesised that catastrophising and internal pain control would mediate disability change. The group discussion condition was to provide an attention-control condition. All potential treatment mediators were assessed via subscales of the Pain Coping and Cognition List: catastrophising, pain coping, internal pain control and external pain control. Residualised change scores for each variable were used in the analysis and correlations were conducted to check that change in the potential mediators was related to change in outcome (a detailed explanation of residualised change is given in Chapter 5). The authors found that change in catastrophising did partially mediate change in activity tolerance, which was used as a measure of disability. However, in this study no mediating effect was calculated, and the treatment groups were collapsed together for the analysis because the authors found no differences between the treatment groups for any potential mediator. They also acknowledge that the two treatment groups both received the same 'base' treatment which focused on the same principles, which may explain why no difference was found between the two active interventions.

Nicassio et al (1997) compared a behavioural intervention consisting of education, relaxation training, goal-setting and prompting with an educational/control intervention involving lectures and group discussion in 86 patients with fibromyalgia. Both conditions were implemented over a 10-week period. Disability was assessed via the Quality of Well-Being Scale, and three potential treatment mediators were tested: helplessness, pain coping and social support. The authors did compute change scores for all included variables and found that change in each potential treatment mediator was correlated with change in outcome. However, the results of the mediation analysis were more difficult to interpret. After conducting regression analyses using the

change scores to assess each potential mediator variable as a predictor of change in outcome, the authors examined “low” and “high” change in helplessness as mediators of depression and pain. They found that change in helplessness and coping was not associated with improvement in disability, so this outcome was not tested further. Change in helplessness was reported as a mediator of change in pain and depression. However, although the authors reported this as a test of mediation, it is not clear exactly how this was conducted and no test of the mediating effect was given.

Two papers used methods that were not previously identified as methods of mediation analysis. Focht et al (2005) used analysis of co-variance (ANCOVA) to test self-efficacy as a mediator of the effects of two separate interventions (dietary weight loss and exercise) in terms of improving mobility in knee OA patients (*n*=252). Both interventions were implemented over an 18-month period. In this study the weight loss intervention involved psychological principles, such as goal setting, problem solving and cognitive restructuring. The authors do report that self-efficacy mediated mobility, but it is unclear whether using ANCOVA is an appropriate way of testing for this and how this tests mediation. The authors reported that they first tested whether changes in self-efficacy and pain independently affected outcome and found that they were both significant predictors. They then tested the effect of the intervention on change in outcome after controlling for change in self-efficacy and pain. They state that this latter analysis resulted in a reduction of the main effect for stair-climbing self-efficacy and pain, but not for walking self-efficacy and pain. This method seems to be similar to the causal steps approach, but the Baron & Kenny (1986) paper is not referred to by the authors.

Finally, Seymour et al (2009) investigated a behaviour change intervention using either physiotherapists or certified exercise instructors, to explore whether the instructor had an impact

on hip/knee OA patient outcomes. 351 patients were recruited and the intervention was carried out over eight weeks. The authors looked at self-efficacy, exercise adherence, attendance and exercise maintenance as potential mediators and stated that self-efficacy was a mediator for both types of instruction post-treatment (two months after the baseline assessment), although not by follow-up (six months after the baseline assessment). However, no method of mediation analysis was formally stated. A random effects model was described in the methods and results, giving values for the main effect and interaction, but not stating any mediating effects.

4.6 Discussion

4.6.1 Summary of psychological constructs identified in previous MSK pain mediation analyses

This review aimed to examine the evidence for psychosocial factors that have been tested as treatment mediators in MSK pain populations, and to examine what methods had been used to conduct this analysis. Ten papers were found to meet the criteria for the review, covering a wide range of populations, MSK conditions and interventions. Several factors were found to mediate the effect of treatment on measures of disability across the studies; reduced helplessness, reduced passive coping, reduced catastrophising, reduced fear-avoidance, reduced psychological inflexibility, improved self-efficacy and a reduction in negative pain beliefs and perceived harmfulness of activity, with the most evidence available for catastrophising, pain coping and self-efficacy. However, the range of different methods used to carry out the mediation analysis in the different papers makes it difficult to ascertain the strength of any of these constructs in terms of their mediating effect. For example, of the four studies that investigated self-efficacy, one used ANCOVA, one used a random effects model and two used linear regression to analyse the mediating effect, all of which give different statistics to interpret. Also, the quality of the

mediation analysis used to identify these factors was generally poor with only two studies conducting appropriate analyses, making it difficult to draw strong conclusions. There were also issues with the design of some of the interventions or aspects of the mediation analysis that could have impacted on the findings of the studies. The key issues are outlined below.

As suggested in Chapter 3, using a theoretical basis for any intervention will give support to the hypothesised temporal order of the variables when conducting the analysis, help researchers to choose appropriate variables to test as mediators and provide evidence to support or refute theoretical models about how an intervention works. Wicksell et al (2010) and Wicksell et al (2013) were the only included papers to use an intervention that was explicitly based on a theoretical model (ACT). Other studies (Smeets et al 2006; Spinhoven et al 2004; Turner et al 2007; Leeuw et al 2008; Robinson et al 2013) included CBT as an intervention which is based on principles of cognitive-behavioural theory, but the papers did not explain how the intervention related to this theory. The remaining studies (Focht et al 2005; Nicassio et al 1997) cited a theoretical framework but did not provide information on how it was used in the intervention, and one study (Seymour et al 2009) did not specify any theoretical basis for their intervention.

Evidence for the psychometric properties of the measures of outcomes and potential mediators was inconsistent across the studies. This is unfortunate as the evidence gathered in Chapter 3 highlighted the importance of only using well validated measures of constructs, as the analysis methods used to carry out mediation analysis often assume there to be no measurement error (see Chapter 3). Information about the characteristics of measures used and whether they have been validated in the population being studied is crucial to assess and report.

In order for a variable to be considered as a potential mediator it needs to be shown to be modifiable and to be shown to be related to change in outcome (MacKinnon 1994). Among the studies included in this review, evidence of such preliminary analysis was inconsistent with most

studies reporting only tests for one or the other of these criteria. Performing such analysis prior to the more complex tests for mediation would be beneficial in terms of helping to decide which variables to focus on, and could possibly be done outside of the confines of a trial, for example in a cohort study (see Chapter 3).

Currently, the evidence suggests that the product of coefficients approach carried out using SEM and obtaining bootstrapped CIs to give an idea of the precision of the indirect (mediating) effect is the most appropriate way to perform mediation analysis. In the present review only four of the 10 included papers conducted their analysis in this way (Turner et al 2007; Wicksell et al 2010; Wicksell et al 2013; Robinson et al 2013) and only Turner et al (2007) reported using SEM to perform their analysis. Further work which performs mediation analysis according to these criteria would help provide more robust evidence for mediating factors.

As explained in Section 4.4.5 (and in more detail in Chapter 3), even in RCTs there is potential for the relationship between the potential mediator and outcome to be affected by unmeasured or confounding factors. Only four studies (Focht et al 2005; Smeets et al 2006; Leeuw et al 2008; Robinson et al 2013) explicitly reported that they had adjusted for confounding factors. A theoretical framework would be helpful in terms of identifying potential factors to adjust for or include in future mediation analysis studies.

A number of authors (e.g. Kazdin 2007) (see Chapter 3) have suggested that at least three assessments of outcome and potential mediators should be carried out in mediation studies, and where such studies involve an intervention assessments need to be conducted during the intervention as well as pre-and post-intervention. This is so that temporality can start to be established. In this review, the majority of studies conducted assessments at three or more time

points but no study conducted assessments while the intervention was taking place. This is a key area for improvement in future mediation studies.

A final point is that none of the studies included in this review provided a power calculation to show that their sample size allowed them to detect a mediating effect. As discussed in Chapter 3, guidance is available on calculating power in mediation studies (Fritz & MacKinnon 2007; Thoemmes et al 2010) which shows that for studies using the Baron & Kenny causal steps approach a sample of more than 20,000 is required to detect an effect of 0.8, while studies using the product of coefficients approach require a sample of 500 to detect the same effect size (Fritz & MacKinnon 2007). The bias-corrected bootstrap has been found to be the most powerful test of the size of the mediated effect (Fritz & MacKinnon 2007; Hayes & Scharkow 2013), although this difference between tests in terms of power decreases as sample size increases (Hayes & Scharkow 2013). The largest study included in this review recruited 351 patients (Seymour et al 2009); most had used the Baron & Kenny approach (of which two had included bootstrapped CIs) and three had used the product of coefficients approach (all of which included bootstrapped CIs). It is acknowledged that recruiting large sample sizes for trials is difficult and that funding for larger trials may be difficult to obtain, which highlights the importance of choosing an appropriate mediation approach so that an effect can be obtained (if it exists) with the smallest possible sample size.

4.6.2 Comparison with existing literature

Systematic reviews in other areas (Cerin et al 2009; Lubans et al 2008; Stanton et al 2013) also failed to draw clear conclusions as to what factors mediated outcome. However, these reviews stated that there was evidence for self-efficacy being a mediator across different interventions. The present review provides some support for these previous findings with three out of the four studies testing self-efficacy reporting that it was a mediator of the relationship between

treatment and disability, suggesting that mediating factors may not be specific to certain fields. The papers included in the present review often tested a range of mediators within their studies.

In terms of the quality of the mediation analysis, the studies included in this review are also similar to those included in the Cerin et al (2009) and Lubans et al (2008) reviews. The criteria added by Cerin et al (2009) which asked whether changes in the mediator preceded changes in outcome was not reported in any paper included in either the present review or Cerin et al's (2009) paper. This point has been noted by other authors in the mediation field (e.g. Kazdin 2009) as being a key area for future research. Being able to establish whether the mediator changes before the outcome is crucial in terms of establishing whether the link between the two is likely to be causal, rather than simply an association. The studies included in the previous reviews were more often based on a specific theoretical framework than those included in the present review. This is also an important area for future research – theory-based interventions allow specific hypotheses to be tested and help give evidence to support or refute current theoretical models (Michie et al 2008). Mediation analysis can also help to refine theory by identifying which aspects are most important to focus intervention efforts on.

4.6.3 Overview of methods of mediation analysis used

One of the objectives of this review was to identify which methods of mediation analysis had been used to conduct mediation analysis in MSK pain populations and explore whether the method used may have impacted on the study findings. It was not possible to identify from the studies included here whether the method used had an impact on their results as even though several studies had used the same method of analysis, each study tested different variables as potential mediators and included assessments at different time points. Chapter 3 described a number of different methods of conducting mediation analysis and outlined the problems with

each. It is clear from the evidence gathered for this systematic review that the suggestions for improvement in methods and the problems identified by experts in the mediation analysis literature are not yet recognised in the applied literature. The regression analysis procedure outlined by Baron & Kenny (1986) was by far the most popular method of mediation analysis used within these studies, despite the limitations of this method being widely cited. Some authors did correct for some of the shortcomings of this approach, for example, by including bootstrapped CIs (Turner et al 2007; Wicksell et al 2010), but problems were still identified in every paper, meaning that confidence in these results may be limited.

4.6.4 Potential limitations of the systematic review

This review was intentionally very broad because of the expected paucity of studies that would have conducted mediation analysis in primary care LBP populations. This has caused complications in terms of being able to compare the studies, but does highlight the need for further research in this area especially in terms of improving the quality of the mediation analysis conducted. There were also some problems with the search strategy in terms of making it sensitive enough to pick up relevant papers but specific enough to ensure appropriate material was found. Methodological terms were included specifically in the search strategy which could have led to only papers that used and described these specific methods in the title or abstract being included. However, extensive citation and reference checking and contact with experts in the literature suggested that no large or influential studies investigating treatment mediation in patients with MSK pain were missed.

The critical appraisal tool used to judge specific aspects of mediation analysis in each of the included papers may not have provided enough information to objectively assess the studies. Each item was only scored according to whether or not the paper provided the necessary

information, and it could be that certain aspects, such as psychometric properties of measures and the appropriateness of the mediation analysis were simply not reported in sufficient detail. Lubans et al (2008) who used their own version of this tool also reported problems in terms of assessing whether or not a study had conducted pilot trials, and that assessing Cronbach's alpha was only one, limited way of assessing the psychometric properties of included measures. The cut-off for Cronbach's alpha suggested by this tool, 0.60, is also lower than other reported cut-offs (for example Terwee et al (2007) recommend a cut-off of 0.70 as a minimum value). However, the items assessed are all aspects of mediation study design and analysis that are reported in the wider literature to be crucial in terms of providing robust evidence for mediation. It is therefore important that future studies of mediation analysis in MSK pain populations specifically report these aspects so that the quality of the analysis can be accurately assessed.

4.7 Conclusion

This review has shown that, although several factors (reduction in helplessness, passive coping, catastrophising, fear-avoidance beliefs, psychological inflexibility, increase in self-efficacy and a reduction in negative pain beliefs and perceived harmfulness of activity) have been found to mediate disability outcomes in MSK pain intervention studies, the quality of the mediation analysis is too weak to draw firm conclusions as to the evidence for psychological mediators of treatment effect on disability outcomes. It also highlights the paucity of studies that have carried out mediation analysis in MSK and LBP intervention studies.

The literature review in Chapter 3 identified a number of areas where future research can be improved. This review shows that in MSK pain mediation studies in particular, further work is needed to improve the quality of the mediation analysis conducted and to design intervention

studies with strong proposed underlying mechanisms that are appropriate for exploring mediating factors, before robust evidence for psychological mediators can be collated. Specifically, future studies should specify a theoretical basis for the intervention, use reliable, previously tested measures and conduct mediation analysis to the standard currently seen as appropriate (i.e. using the product of coefficients approach and including bootstrapped CIs of the indirect (mediating) effect).

Another aspect identified in Chapter 3 was including at least three assessment points. In this review, five of the ten studies included at least three assessment points but these assessments all occurred at baseline and then at post-treatment or follow-up. One of the key recommendations from the literature was that assessments of change in outcome and change in the potential mediators should be made during treatment to help establish issues around causality and temporality (see Chapter 3). This should be a focus of future mediation research in MSK pain populations.

A final point that was identified in Chapter 3 and also highlighted as a key issue at a workshop on design aspects of mediation analysis at the Low Back Pain Forum XII conference in 2012 (Mansell et al 2012) is the concept of mediation analysis as a programme of work, starting with qualitative, exploratory methods to identify key mediating factors; using cohort study data to test the preliminary mediation steps of whether or not a factor is modifiable and whether change in the potential mediator is related to change in outcome; and only then moving on to test those factors most likely to be mediators of treatment outcome in intervention studies. No study included in this review reported conducting pilot studies prior to their intervention study, but such preliminary work would be valuable in helping researchers get the most out of mediation analysis nested within intervention studies.

The next step in this thesis is to apply the appropriate mediation analysis methods to cohort study data and RCTs of interventions to improve functioning in MSK pain patients, to add to the evidence base in this area. Although this analysis will not be able to address all of the issues identified in this review and in Chapter 3, the issues pertaining to the analysis will be focused on as an area for improvement. The analysis will include several of the factors found by the review to be mediators of disability measures.

Chapter 5: Using Observational Data to Identify Potential Treatment Mediators of Functional Outcome in a Primary Care Musculoskeletal Pain Population

5.1 Introduction

Chapter 3 discussed design issues of conducting mediation analysis and concluded that identifying treatment mediators involves a process of research which may also include observational studies to identify potential mediation constructs. This chapter will therefore focus on mediation research conducted in observational studies of patients with musculoskeletal (MSK) pain, and will also provide an example of the use of observational data to explore potential mediating factors in primary care consulters with low back pain (LBP), in preparation for treatment mediation analysis using trial data reported in Chapter 6. In this thesis, 'observational studies' are studies that employ a cross-sectional or longitudinal design and participants may receive any type of intervention in the context of usual care, rather than a specific intervention as part of the study. This means that mediation analysis does not concern mediators of a specific treatment effect but of the association between two key variables of interest. While such observational studies cannot establish which variables are treatment mediators, they can provide important information on which variables change over time and which are related to the outcome of interest, and help build evidence for particular factors as possible mediators of treatment outcome.

5.1.1 Mediation analysis in observational studies

This section outlines the evidence for mediating factors in MSK pain populations that has been previously conducted. This is not an exhaustive review, but instead provides examples of previously published mediation analyses using observational data and the different analysis

techniques and study designs employed.

The majority of studies in MSK pain populations conducting mediation analysis in its broadest sense (i.e. examining mediators of treatment or other (causal) pathways) have used a cross-sectional or cohort design. The participants in these studies were mostly recruited from secondary or tertiary care pain clinics or physiotherapy departments, although the studies themselves did not provide any specific form of treatment. In these types of studies, the construct of self-efficacy has been commonly investigated for its role as a mediator in pathways to functional outcomes. For example, this construct was tested as a mediator of the relationship between pain and disability (Arnstein et al 1999; 2000; Söderlund & Åsenlöf 2010; Costa et al 2011), between pain-related fear and disability (Woby et al 2007), between pain catastrophising and disability (Shelby et al 2008), and between resilience and disability (Johnson Wright et al 2008). Other variables investigated as potential mediators include acceptance as a mediator of the relationship between catastrophising and disability (Vowles et al 2008), pain beliefs and attitudes as a mediator between pain and disability (Alcantara et al 2010), anxiety and depression as a mediator between pain and disability (Arnstein 2000; Wegener et al 2011), pain catastrophising as a mediator between pain and disability (George et al 2011), and fear-avoidance beliefs as a mediator between catastrophising and disability (Leeuw et al 2007b) and between self-efficacy or conflict and disability (Karoly et al 2008). However, often only one study has tested a particular psychological factor in a particular pathway, and many of these different psychological constructs are likely to overlap, making it difficult to draw conclusions as to the true mediating effects of these variables on outcome and how they inter-relate to explain change in disability. Most recently, a systematic review and meta-analysis of observational studies summarised the results of 12 studies investigating mediators of the relationship between pain and disability (Lee et al 2015), reporting evidence for pain self-efficacy, psychological distress and fear-avoidance beliefs

as potential mediators of treatment effect.

There are some difficulties when testing mediators in observational data, especially when using cross-sectional data. For example, Costa et al (2011) tested self-efficacy as a mediator of the association between pain and disability, and compared this with fear-avoidance beliefs. These variables were chosen to help test two theoretical models: social learning theory (Bandura 1977), which suggests confidence in one's ability to do things is related to one's ability to achieve goals; and the fear-avoidance model (Vlaeyen & Linton 2000), which suggests that people who experience pain and perceive it to be as a result of injury may be afraid to move in case they cause further damage, leading to disuse and disability. This study involved data at two time points, allowing the authors to examine mediation in terms of change as well as cross-sectionally using linear regression analysis. The sample involved patients with a back pain episode of more than three months duration (*n*=184). Interestingly, the authors reported that both self-efficacy and fear-avoidance beliefs were found to be mediators of the association between pain and disability when analysing cross-sectional baseline data, but when mediation analysis was conducted using change in the potential mediators and change in disability (from baseline to 12-month follow-up), only change in self-efficacy (not fear-avoidance beliefs) was found to partly explain changes in function. The authors do not offer a potential explanation for this finding but it may be due to fear-avoidance beliefs changing less between baseline and 12-months than self-efficacy (mean values for fear-avoidance beliefs 40.5 at baseline and 39.1 at follow-up; self-efficacy 44.4 at baseline and 49.0 at follow-up). If there is no or little change over time in a variable, it is unlikely to be mediating any change seen in the outcome. More variance in the outcome was explained in the cross-sectional models than the longitudinal models (53% for baseline self-efficacy and 34% for change in self-efficacy), which suggests other factors may be affecting outcome over time. This shows how potentially different results can be obtained when

follow-up points are included in a study; cross-sectional associations may appear to be stronger because they do not take into account how variables may change over time. Temporality (the order in which variables change), an issue discussed in detail in Chapter 3, is also very important in mediation analyses which explore a causal association between the mediator and outcome. Results of cross-sectional studies should therefore be interpreted with caution as these studies do not provide evidence of temporality and associations are likely to be weaker when investigated in longitudinal or intervention study designs.

The statistical approaches used to test for mediation have also differed across previous mediation studies in this field, which could have influenced the results found. The majority of studies employed linear regression, often citing Baron & Kenny's causal steps approach (Arnstein et al 1999; Costa et al 2011; Woby et al 2007; Leeuw et al 2007b; George et al 2011; Söderlund & Åsenlöf 2010; Alcantara et al 2010). However, several problems with using this method have been identified (see Chapter 3). For example, in the cross-sectional study by Leeuw et al fear-avoidance beliefs were tested as a mediator of the relationship between catastrophising thoughts and disability as proposed by the fear-avoidance model (Vlaeyen & Linton 2000). Fear-avoidance beliefs did not meet the first criterion of the Baron & Kenny (1986) regression analyses, with catastrophising not having a direct effect on disability, so no further analysis was performed using this variable. It has been suggested that this first step is not always necessary to perform mediation analysis, and indeed requiring this can reduce the power to detect mediation (MacKinnon et al 2007 – see Chapter 3). Many of these studies employed an additional test of the mediating effect but the majority used the Sobel (1982) test which assumes the effect is normally distributed. This could potentially mean that the results presented by these studies do not accurately represent the mediating effects of the tested variables. Only two studies (Shelby et al 2008; Vowles et al 2008) employed a test of the mediating effect that is seen as appropriate by

key authors in the mediation analysis field (e.g. Hayes 2009; MacKinnon 2004) (see Chapter 3), where the likely non-normal distribution of the indirect effect is accounted for by the use of bootstrapping. The sample sizes included by the studies also appeared to be small ($n \leq 100$ in some cases (George et al 2011)) which could also impact on the power of the analysis to detect mediating effects. In their systematic review, Lee et al (2015) also noted limitations of the included studies, including a lack of consideration of potential confounding factors, use of inappropriate mediation analysis techniques and a lack of consideration or testing of the temporal order of the variables.

However, a strength of the observational mediation research discussed above is that almost all based their analysis on a strong theoretical model, in particular Vlaeyen & Linton's fear-avoidance model (Johnson Wright et al 2008; George et al 2011) and Bandura's model of social-cognitive theory (Arnstein 2000; Arnstein et al 1999; Woby et al 2007). Even though most of the research described above is cross-sectional, a theory or hypothesis behind why variables are linked in a particular way is helpful for exploring associations and building an evidence base for potential mediators to be tested in further research.

To summarise, the majority of the previous studies of mediation analysis in MSK pain populations have carried out the analysis on observational data. This has provided some support for theoretical models of pain, but the methods were often flawed and few variables had been extensively tested. The work presented in this chapter used observational data for a different purpose; to identify what factors could be potential treatment mediators and should therefore be tested in a randomised controlled trial (RCT).

5.2 Aim

The overall aim of this thesis was to explore methodological issues in conducting mediation analysis in MSK pain intervention studies. By using observational study data to decide which variables to test as treatment mediators in RCT data, this chapter demonstrates how the investigation of mediators involves a process of research to build up evidence for particular factors before they are tested as mediators of treatment effect. The purpose of such studies is to allow the selection of variables for mediation analysis that are likely to show change and are related to change in outcome.

The aim of the secondary analysis presented in this chapter was to investigate which psychological factors are likely to contribute to change in functional outcome, and might therefore be useful to test as treatment mediators in an RCT. To answer this, associations were investigated between changes in psychological predictor variables and changes in outcome between baseline and each follow-up point. The study which was used for this analysis, the BeBACK study, is described in more detail in Section 5.5.

5.3 Specific Objectives

1. To provide descriptive data on the outcome (disability) and each of the predictor (potential mediator) variables at baseline and each follow-up point (three months, six months and 12 months);
2. To explore the relationships between change in each of the predictor (potential mediator) variables and change in the outcome measure (disability).

5.4 Hypothesis

Change in the following prognostic variables: self-efficacy, measures of coping (catastrophising thoughts, diversion, re-interpretation, cognitive coping), fear-avoidance beliefs, anxiety, depression and illness perceptions (identity, personal control, acute-chronic timeline, consequences, timeline cyclical, illness coherence, treatment control, emotional representation, psychological attribution, risk factors, immunity and accident/chance) is significantly related to change in the outcome variable (disability) at three-month, six-month and 12-month follow-up.

5.5 Beliefs about Back Pain (BeBACK) Study

The BeBACK study is a prospective, observational cohort study of adult patients consulting their GP with LBP. The purpose of the original study was to investigate which of 20 psychological factors, previously found to be associated with LBP, were related to LBP at presentation in the present sample and also which were predictive of changes in disability (Foster et al 2010). The Roland-Morris Disability Questionnaire (RMDQ) (Roland & Morris 1983) and a self-reported global rating of change were used as primary outcome measures. The main prognostic variables included in the analyses were illness perceptions as measured by the Illness Perceptions Questionnaire Revised (IPQ-R) (Moss-Morris et al 2002), as these were the main focus of the study (Foster et al 2008). However, other measures, including self-efficacy, impact and severity of pain, coping strategies, anxiety, depression, bothersomeness, fear-avoidance beliefs and pain severity were also included, along with demographic and work information. Foster et al (2010) found that out of the 20 measures tested in a multivariate regression analysis, only four were found to be independently predictive of disability and global rating of change at six-month follow-up. These were self-efficacy and three illness perceptions (personal control, acute/chronic timeline and

identity). In the original study, the main aim of the analysis was to test all of the variables as predictors of longer-term outcome, rather than as potential mediators, so associations between baseline values of the psychological factors with follow-up assessments of outcome were tested.

However, Foster et al (2010) also reported that they explored the association between change in each of the psychological constructs and change in outcome, and found statistically significant associations for eight out of the 20 constructs of which change in only three were independently predictive of change in disability (illness identity, pain self-efficacy and depression), which explained 42.4% of the variance of change in disability. These variables were therefore suggested to be potential mediators of disability outcome, although the authors acknowledged that this could not be assumed to indicate a causal relationship. This analysis did not control for baseline level of the outcome measure (disability), which is generally highly predictive of outcome. Further analysis which controls for this potential confounding factor is therefore warranted.

Further analysis of the BeBACK study was undertaken to investigate whether the prognostic factors identified previously were also potential treatment targets (mediating factors). This study has a number of strengths in that it is sufficiently large, includes several time points (three, six and 12 months follow-up), and measured a large number of psychological variables that can be examined for potential mediating effects. However, unlike in the studies discussed in Section 5.1.1, no formal test of mediation analysis was carried out on the BeBACK data, as the variables in this study were only be tested for their *potential* to mediate treatment outcome. The purpose of this analysis was to explore which variables changed over time and which were related to change in disability, the outcome of interest.

The analysis presented here includes the RMDQ as an outcome variable at three-, six- and 12-month follow-up. This is comprised of a list of 24 statements related to the ability to carry out movements or everyday activities, which the respondent checks if they feel that particular statement applies to them on the day they complete the questionnaire. A higher score therefore indicates higher back-pain-related disability.

All of the psychological variables included in the original study were included in the present analysis as potential mediating variables (Table 5.1). These were self-efficacy (Pain Self-Efficacy Questionnaire (PSEQ)) (Nicholas 2007), illness perceptions (Illness Perceptions Questionnaire – Revised (IPQ-R)) (Moss-Morris et al 2000), fear-avoidance beliefs (Tampa Scale of Kinesiophobia (TSK)) (Kori et al 1990), pain catastrophising and coping (Coping Strategies Questionnaire-24 (CSQ-24)) (Harland & Georgieff 2003), and anxiety and depression (Hospital Anxiety and Depression Scale (HADS)) (Zigmond & Snaith 1983).

The PSEQ contains a list of 10 items that are scored on a six-point Likert scale. Respondents rate each item in terms of how confident they are in their ability to do things such as socialising and household chores despite their pain. The CSQ contains four subscales for catastrophising, diversion, re-interpretation and cognitive coping, all of which are scored on a seven-point Likert scale to indicate how often a particular coping strategy is used. This scale is a shortened version of the original Coping Strategies Questionnaire developed by Rosenstiel & Keefe (1983). The TSK is made up of 17 items which are each rated on a four-point Likert scale, with a higher score indicating higher fear-avoidance beliefs; a score of 37 or above has been documented as a cut-off for high fear (Vlaeyen et al 1995a; Bränström & Fahlström 2008) and a reduction in score by four points as showing clinically significant improvement (Woby et al 2005). The HADS contains 14 items, with seven items each for anxiety and depression. These are also scored on a four-point

Likert scale, with a higher score indicating higher anxiety and/or depression. Finally, the IPQ-R is also made up of a number of subscales which assess various illness perceptions and is designed so that the statements can be made specific to the illness being studied. The Identity subscale requires respondents to answer 'yes' or 'no' to whether they think a number of different symptoms were related to their illness. The items of the remaining subscales (Personal Control, Acute-Chronic Timeline, Consequences, Timeline Cyclical, Illness Coherence, Treatment Control, Emotional Representation, Psychological Attribution, Risk Factor, Immunity, and Accident/Chance) are scored using five-point Likert scales which require respondents to rate how strongly they agree or disagree with certain statements. A summary of each of the measures is provided in Table 5.1 below.

Table 5.1 Description of outcome and potential mediator measures

Questionnaire/Sub-scale	Reference	How scored
Outcome measure		
Disability: Roland-Morris Disability Questionnaire (RMDQ)	Roland & Morris (1983)	Add up number of 'Yes' rated responses – higher number of 'Yes' responses indicates higher level of disability (score of 0-24)
Potential mediators		
Self-efficacy: Pain Self-efficacy Questionnaire (PSEQ)	Nicholas (2007)	Likert scale to record responses – higher scores indicate higher self-efficacy
Coping: Coping Strategies Questionnaire (CSQ24): Catastrophising subscale	Harland & Georgieff (2003)	6 items on a 7 point scale (0=never use it; 6=always); score of 0-36; Higher score indicates a higher frequency of the coping style
Coping: CSQ24: Diversion Subscale	Harland & Georgieff (2003)	6 items on a 7 point scale (0=never use it; 6=always); score of 0-36; Higher score indicates a higher frequency of the coping style
Coping: CSQ24: Re-interpretation subscale	Harland & Georgieff (2003)	6 items on a 7 point scale (0=never use it; 6=always); score of 0-36; Higher score indicates a higher frequency of the coping style
Coping: CSQ24: Cognitive coping subscale	Harland & Georgieff (2003)	5 items on a 7 point scale (0=never use it; 6=always); score of 0-30; Higher score indicates a higher frequency of the coping style
Fear-avoidance beliefs: Tampa Scale for Kinesiophobia (TSK)	Kori et al (1990)	17 items on a 4-point Likert scale "strongly agree"-"strongly disagree"; Higher score indicates higher fear-avoidance
Anxiety: Hospital Anxiety and Depression Scale (HADS)	Zigmond & Snaith (1983)	7 items on a 4 point scale (0-3); Higher score indicates higher level of anxiety
Depression: HADS	Zigmond & Snaith (1983)	7 items on a 4 point scale (0-3); Higher score indicates higher level of depression
Illness Perceptions: Illness Perceptions Questionnaire-Revised (IPQ-R): Identity subscale	Moss-Morris et al (2002)	14 items, each answered either "yes" or "no"; Sum all "Yes" rated responses to each commonly experienced symptom – Higher score indicates stronger beliefs about the number of symptoms attributed to the illness
Illness Perceptions: IPQ-R: Personal Control subscale	Moss-Morris et al (2002)	6 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree) Add up IP12-IP17 (Section C) – higher score indicates more positive beliefs about controllability of illness
Illness Perceptions: IPQ-R: Acute-Chronic Timeline subscale	Moss-Morris et al (2002)	6 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree) Add up IP1-IP5 + IP18 (Section C); Higher score indicates stronger beliefs about the chronicity of the illness

Questionnaire/Sub-scale	Reference	How scored
Illness Perceptions: IPQ-R: Consequences subscale	Moss-Morris et al (2002)	6 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items IP6-IP11 (Section C); Higher score indicates more of an effect of the illness on the patient
Illness Perceptions: IPQ-R: Timeline Cyclical subscale	Moss-Morris et al (2002)	4 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items IP29-IP32 (Section C); Higher score indicates stronger beliefs about the cyclical nature of the condition
Illness Perceptions: IPQ-R: Illness coherence subscale	Moss-Morris et al (2002)	5 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items IP24-IP28 (Section C); Higher score indicates positive personal understanding of condition
Illness Perceptions: IPQ-R: Treatment Control subscale	Moss-Morris et al (2002)	5 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items IP19-IP23 (Section C); Higher score indicates higher beliefs in treatment control
Illness Perceptions: IPQ-R: Emotional representation subscale	Moss-Morris et al (2002)	6 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items IP33-IP38 (Section C); Higher score indicates more negative emotion as a result of the illness
Illness Perceptions: IPQ-R: Psychological attribution subscale	Moss-Morris et al (2002)	6 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items 1, 9, 10, 11, 12 and 17 (Section D); Higher score indicates higher number of psychological attributions
Illness Perceptions: IPQ-R: Risk factor subscale	Moss-Morris et al (2002)	7 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items 2, 4, 6, 8, 13, 14 and 15 (Section D); Higher score indicates higher number of perceived risk factors
Illness Perceptions: IPQ-R: Immunity subscale	Moss-Morris et al (2002)	3 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items 3, 7 and 18 (Section D); Higher score indicates perceived poor immunity
Illness Perceptions: IPQ-R: Accident/Chance subscale	Moss-Morris et al (2002)	2 items on a 5-point Likert scale (Strongly Agree to Strongly Disagree); Add up items 5 and 16 (Section D); Higher score indicates higher attribution to chance

Shaded subscales: Items can be modified to suit particular illnesses (Moss-Morris et al 2002)

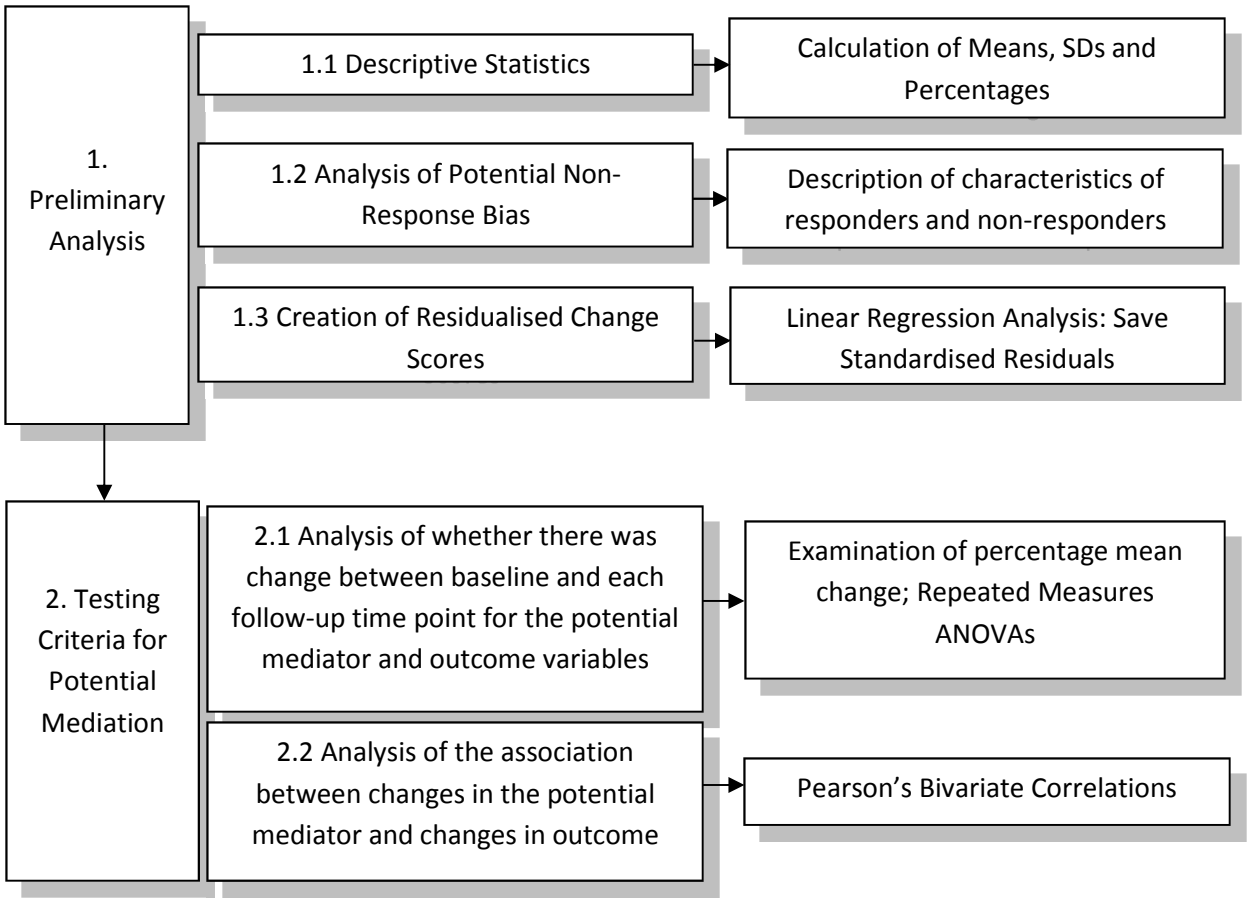
The psychometric properties of these measures have been reviewed. The IPQ-R was found to have good internal consistency and test-retest reliability by the authors who originally re-developed the tool (Moss-Morris et al 2002; Hill 2010) although these properties were not tested specifically in a LBP population. The PSEQ has been tested in a LBP population and was found to have strong internal consistency, test-retest reliability and unidimensionality (Nicholas 2007; van Hartingsveld et al 2010). The CSQ-24 has been reported to have good internal consistency and a consistent factor structure in MSK pain populations (Harland & Georgieff 2003; Harland & Martin 2014). A systematic review of the psychometric properties of the TSK located very few studies that had investigated the 17-item version of the tool – several other versions exist including a 13-item and 11-item version (Lundberg et al 2011). The authors of this review noted that none of the studies included had measured the responsiveness of the TSK, and that overall validity was low (Lundberg et al 2011). Studies have also cast doubt on the factor structure of the TSK, with a two-factor structure (Goubert et al 2004; French et al 2007) and a five-factor structure found previously (Lundberg et al 2004). However, this evidence is not consistent and several authors found poor fit for the two-factor structure (see Tkachuk & Harris 2012 for a review). Others have found support for the TSK being a unidimensional construct (Damsgård et al 2007). For the HADS, less evidence exists for the tool's psychometric properties in MSK pain populations, and this measure was originally developed for a secondary care rather than a primary care population. Internal consistency (Cronbach's alpha) for the HADS in mental health populations has been reported as 0.76 and 0.80 for the depression and anxiety subscales respectively (Mykleton et al 2001) and a review by Bjelland et al (2002) noted good sensitivity and specificity in primary care populations. However, recently there have been calls to abandon use of the HADS over its inability to distinguish between anxiety and depression (Coyne & van Sonderen 2012). The tool's power to predict LBP and disability in primary care studies is mixed, with some suggesting the tool is predictive (Foster et al 2010) and others suggesting not (Campbell et al 2013a). Finally, the 24-

item RMDQ has been found to have good psychometric properties overall in LBP populations, with studies indicating adequate construct (Roland & Fairbank 2000; Grotle et al 2005) and content (Roland & Morris 1983; Roland & Fairbank 2000; Ostelo & de Vet 2005) validity, good test-retest reliability over short time periods (Roland & Morris 1983; Streiner & Norman 2008; Smeets et al 2011) and high internal consistency (Roland & Fairbank 2000). Responsiveness in terms of minimal clinically important change (MCIC) yields different results within different populations (Bombardier et al 2001), although Jordan et al (2006) found that a 30% reduction in score is indicative of clinically important change in a primary care population. It should be noted that clinically important absolute change is highly dependent on baseline RMDQ score (Cleland et al 2010).

5.6 Methodology

All analyses were carried out using SPSS PASW statistics package version 18. A flowchart summarising this analysis can be seen in Figure 5.1.

Figure 5.1 Flowchart describing the steps of analysis



5.6.1 Preliminary analyses (Objective 1)

Descriptive statistics (means and standard deviations (SDs)) for numerical measures, or frequencies for binary or categorical measures were calculated for each potential mediator variable and the outcome measure at baseline and each follow-up point (Step 1.1 in Figure 5.1). This was to see if the measures showed a pattern of change over time.

Analysis to detect potential attrition (non-response) bias due to selective loss to follow-up was conducted by examining differences in baseline characteristics between participants responding and not responding at six-month and 12-month follow-up (Step 1.2). This was important given the high dropout rate over the study's follow-up periods. Comparisons between responders and non-

responders were made for age, sex, baseline disability score, and baseline scores on each of the potential mediator variables.

Residualised change scores were then calculated for each potential mediator and outcome measure for each follow-up point (Step 1.3). These scores represent the difference between the score a person receives at follow-up compared to what was predicted at baseline (Streiner & Norman 2008). Raw change scores (follow-up score subtracted from baseline score) such as those used in the original BeBACK analysis do not account for the predictive power of baseline score. This means that residualised change scores are less prone to error than raw change scores (George et al 2008) and give a more accurate picture of the change that actually occurred between the measured time points. Residualised change scores are calculated by running a linear regression analysis with the follow-up score as the outcome, and baseline score as the predictor variable. Residuals (the difference between the observed value at follow-up and the predicted value) are then saved, and it is these values that are then used in the analysis.

5.6.2 Testing criteria for potential mediation (Objectives 2 and 3)

Mean percentage change between baseline and each follow-up mean score for all included measures was calculated and the magnitude of change that occurred at each point was examined. Repeated measures Analysis of Variance (ANOVA) was then conducted to test if change in each of the outcome and potential mediator measures was statistically significant between baseline and follow-up. This is an important requirement for showing a process of change (Vowles et al 2007). ANOVA tests whether there is a difference in group means; here, a repeated measures ANOVA was used because the means are not between two separate groups but rather a comparison between different means measured in the same participants over time. However, because of the large sample size in the BeBACK dataset it is likely that even small changes will be statistically

significant. Therefore, the mean change in each variable, expressed as a percentage, was presented alongside the repeated measures ANOVAs to give more information as to the magnitude of this change. Mean scores at baseline, three-month, six-month and 12-month follow-up for each variable were entered into the repeated measures ANOVA to test whether there was a significant change across all the time points.

The residualised change score of each potential mediator variable was then correlated with the residualised change score of the outcome measure at each time point using Pearson's bivariate correlations, to examine the relationship between change in the potential mediators and change in outcome. Again, showing a relationship between change in the potential mediator and change in outcome is important for establishing whether a variable has the potential to mediate outcome.

5.7 Results

5.7.1 Preliminary analyses (Objective 1)

Participants were sent a baseline survey which asked for consent for further contact. Only those who responded at baseline and then consented to further contact were sent follow-up questionnaires. The baseline population (*n*=1591) contained 41% male and 59% female participants, with a mean age of 43.9 (SD 10.3) years. Pain duration was measured in categories of Less than 1 month, 1-3 months, 4-6 months, 7 months-3 years or More than 3 years, with similar proportions of patients in each category. Of the participants who responded at baseline, 54% (*n*=856) responded at three-month follow-up, 51% (*n*=810) at six-month follow-up and 30% (*n*=473) at 12-month follow-up. The results of the analysis of potential non-response bias is presented below, using means and SDs or percentages where appropriate. Analysis of baseline

characteristics showed that responders at three-month follow-up tended to be slightly older, have higher self-efficacy, catastrophising thoughts, cognitive coping, identity, psychological attribution and risk factor scores and lower treatment control and accident/chance scores compared to non-responders, but these differences appear very small, with often less than one point difference between responders and non-responders. Some of the scales, particularly the illness perceptions subscales, only include a few items but the differences between responders and non-responders for these scales (Timeline Cyclical, Immunity and Accident/Chance) were all less than one point. By six-month follow-up a greater percentage of responders were female but there still appeared to be only small differences between baseline values for responders and non-responders. By 12-month follow-up there was a bigger difference between responders and non-responders for baseline distribution of sex, pain duration of 7 years or more, self-efficacy, cognitive coping and fear-avoidance beliefs in particular, with more females and those with more favourable scores on psychological measures remaining in the study. These results are summarised in Table 5.2.

Table 5.2 Analysis of potential non-response bias in BeBACK participants (baseline characteristics of responders and non-responders to follow-up questionnaires)

Measure		**Baseline values (mean (SD)) or n(%) (*n*=1591)**	**Baseline values (Mean (SD)) at three-month follow-up**		**Baseline values (Mean (SD)) at six-month follow-up**		**Baseline values (Mean (SD)) at 12-month follow-up**	
			Resp (*n*=856)	**Non-resp (*n*=735)**	**Resp (*n*=810)**	**Non-resp (*n*=781)**	**Resp (*n*=473)**	**Non-resp (*n*=1118)**
Age		43.88 (10.31)	45.31 (9.82)	42.23 (10.63)	45.54 (9.75)	42.16 (10.61)	46.23 (9.57)	42.89 (10.46)
Sex		58.5% f	58.8% f	58.1% f	61.1% f	55.7% f	60.9% f	57.4% f
Pain duration	Less than 1 month	288 (18.1)	138 (16.1)	***	129 (15.9)	115 (14.2)	75 (15.9)	62 (13.1)
	1-3 months	329 (20.7)	173 (20.2)	***	156 (19.3)	86 (10.6)	96 (20.3)	54 (11.4)
	3-6 months	243 (15.3)	132 (15.4)	***	134 (16.5)	46 (5.7)	78 (16.5)	36 (7.6)
	7 months -3 years	347 (21.8)	199 (23.2)	***	190 (23.5)	156 (19.3)	107 (22.6)	85 (18.0)
	More than 3 years	338 (21.2)	199 (23.2)	***	186 (23.0)	174 (21.5)	108 (22.8)	118 (24.9)
RMDQ score		8.64 (6.04)	8.76 (5.91)	8.49 (6.19)	8.71 (5.79)	8.57 (6.29)	8.36 (5.82)	8.76 (6.13)
Self-efficacy score		37.81 (14.64)	38.21 (14.33)	37.34 (14.83)	38.43 (14.17)	37.16 (15.00)	40.0 (13.74)	36.88 (14.81)
Coping: Catastrophising score		9.96 (7.87)	9.63 (7.55)	10.34 (8.43)	9.65 (7.60)	10.29 (8.33)	9.38 (7.29)	10.21 (8.24)
Coping: Diversion score		15.51 (8.24)	***	***	15.62 (8.18)	15.38 (8.31)	15.41 (8.40)	15.55 (8.17)
Coping: Re-interpretation score		7.83 (6.99)	***	***	7.59 (6.81)	8.09 (7.18)	7.49 (6.85)	7.98 (7.05)
Coping: Cognitive score		16.27 (6.48)	***	***	16.78 (6.19)	15.72 (6.74)	17.13 (6.01)	15.89 (6.65)
Fear-avoidance beliefs score		39.67 (6.89)	***	***	39.36 (6.86)	40.00 (6.91)	38.89 (6.83)	40.01 (6.89)
Anxiety score		8.26 (4.55)	***	***	8.25 (4.44)	8.28 (4.67)	8.11 (4.37)	8.33 (4.63)
Depression score		6.53 (4.37)	***	***	6.38 (4.15)	6.68 (4.58)	6.15 (4.06)	6.69 (4.48)
Identity score		4.03 (2.38)	***	***	4.11 (2.32)	3.95 (2.43)	4.08 (2.20)	4.01 (2.45)
Personal control score		20.52 (3.79)	20.93 (3.78)	20.02 (3.74)	21.00 (3.79)	20.05 (3.74)	21.27 (3.87)	20.19 (3.71)
Acute-chronic timeline score		19.64 (5.85)	20.02 (5.86)	19.26 (5.81)	19.81 (5.84)	19.46 (5.85)	19.82 (5.81)	19.57 (5.87)
Consequences score		17.33 (5.48)	17.45 (5.48)	17.17 (5.47)	17.34 (5.45)	17.31 (5.51)	17.06 (5.42)	17.44 (5.50)
Timeline cyclical score		13.04 (3.39)	13.07 (3.35)	13.00 (3.45)	13.05 (3.32)	13.03 (3.48)	12.83 (3.33)	13.13 (3.42)

Measure	**Baseline values (mean (SD)) or n(%) (*n*=1591)**	**Baseline values (Mean (SD)) at three-month follow-up**		**Baseline values (Mean (SD)) at six-month follow-up**		**Baseline values (Mean (SD)) at 12-month follow-up**	
		Resp (*n*=856)	**Non-resp (*n*=735)**	**Resp (*n*=810)**	**Non-resp (*n*=781)**	**Resp (*n*=473)**	**Non-resp (*n*=1118)**
Illness coherence score	13.77 (4.99)	13.62 (5.06)	13.94 (4.89)	13.45 (3.10)	14.10 (4.84)	12.99 (5.08)	14.10 (4.91)
Treatment control score	16.99 (3.36)	17.03 (3.37)	16.95 (3.36)	17.17 (3.29)	16.81 (3.43)	17.46 (3.29)	16.80 (3.37)
Emotional representation score	16.71 (5.23)	16.71 (5.23)	16.71 (5.24)	16.68 (5.20)	16.76 (5.27)	16.46 (5.10)	16.82 (5.28)
Psychological attribution score	11.93 (4.11)	***	***	11.73 (4.01)	12.14 (4.21)	11.55 (4.01)	12.09 (4.15)
Risk factor score	15.02 (4.15)	***	***	14.29 (4.01)	15.13 (4.29)	14.84 (4.00)	15.10 (4.21)
Immunity score	5.33 (1.95)	***	***	5.22 (1.90)	5.44 (2.00)	5.13 (1.86)	5.52 (1.98)
Accident/Chance score	5.98 (1.92)	***	***	6.02 (1.93)	5.94 (1.90)	6.01 (1.95)	5.96 (1.90)

*f = female; ***=Not measured*

The above analysis suggests that there are differences between responders and non-responders for some of the measures that will be analysed, but these differences are small and unlikely to strongly influence the results of the associations presented in this chapter. Therefore, despite the high loss to follow-up and loss of precision all of the measures used in the subsequent analysis were based on participants with complete data. Imputation of missing values was not considered for this analysis, as it would involve imputing 50-70% of data.

The skewness and kurtosis values were also examined for each variable and histograms were produced (see Appendix 5.1). This is because all of the analyses conducted in this chapter will assume normal distribution of the variables. This analysis was carried out on both baseline, follow-up values and residualised change scores as all types of scores were used in the subsequent analysis. Some variables did have values of skewness (three-month, six-month and 12-month follow-up mean scores in pain catastrophising; 12-month mean score in disability) or kurtosis (residualised change at three-months in self-efficacy, personal control, pain catastrophising, treatment control; residualised change at six months in fear-avoidance, identity, emotional representation, treatment control, psychological attributions, risk attributions and immunity; residualised change at 12 months in pain catastrophising, depression, personal control, identity, treatment control and immunity) that exceeded 1. Non-normality was also indicated by examination of the histograms, although this was less marked in the residualised change scores (see Appendix 5.1). This indicated that the assumption of normality was not always met, which could have introduced bias into the analysis.

5.7.2 Testing criteria for potential mediation (Objectives 2 and 3)

The means and SDs of the outcome measure and each potential mediator variable for all of the time-points are shown in Table 5.3. The percentage of mean change from baseline to each follow-up time point is also given, as well as the *F* ratio from the repeated measures ANOVA and whether this represented a statistically significant ($p<0.05$) change overall.

Not all variables were measured between baseline and three-month follow-up, but out of those that were, the largest percentage mean change from baseline score was in disability (23.5%) and catastrophising (14%), with very small changes of less than 10% in the other variables measured at this time point.

Table 5.3 Change over time in potential mediator and outcome variables in BeBACK participants

Measure		**Baseline (Mean (SD)) (*n*=1591)**	**Three-month follow-up (Mean (SD)) (*n*=856)**	**Proportion of change from baseline (%)**	**Six-month follow-up (Mean (SD)) (*n*=810)**	**Proportion of change from baseline (%)**	**12-month follow-up (Mean (SD)) (*n*=473)**	**Proportion of change from baseline (%)**	**ANOVA (*F*)**
RMDQ		8.64 (6.04)	6.61 (6.09)	23.5%	6.19 (6.14)	28.4%	5.83 (6.12)	32.5%	71.21**
Self-efficacy		37.81 (14.56)	39.96 (14.11)	5.7%	40.80 (14.15)	7.9%	42.17 (14.07)	11.5%	10.52**
Coping	Catastrophising	9.96 (7.97)	8.50 (7.95)	14.0%	8.70 (7.51)	12.7%	8.19 (7.20)	17.8%	16.15**
	Diversion	15.51 (8.24)	***	***	15.49 (8.71)	0.1%	14.68 (8.42)	5.4%	0.31
	Re-interpretation	7.83 (6.99)	***	***	8.71 (7.29)	11.2%	6.83 (7.00)	10.2%	5.01**
	Cognitive coping	16.27 (6.48)	***	***	17.29 (6.28)	6.3%	17.74 (6.26)	9.0%	2.92
Fear-avoidance beliefs		39.67 (6.89)	***	***	38.38 (6.30)	3.3%	37.80 (6.11)	4.7%	10.94**
Anxiety		8.26 (4.55)	***	***	6.60 (4.57)	20.1%	6.31 (4.48)	23.6%	74.79**
Depression		6.53 (4.37)	***	***	5.08 (4.28)	22.2%	4.67 (4.32)	28.5%	54.76**
Illness perceptions	Personal control	20.51 (3.79)	20.77 (4.17)	1.3%	20.93 (4.16)	2.1%	21.26 (4.27)	3.7%	0.70
	Acute-chronic timeline	19.64 (5.84)	20.92 (5.81)	6.5%	21.00 (5.80)	6.9%	21.38 (5.91)	8.9%	17.40**
	Identity	4.03 (2.37)	***	***	3.14 (2.41)	22.1%	3.02 (2.23)	25.1%	83.32**
	Consequences	17.33 (5.48)	17.06 (5.82)	1.6%	16.49 (5.80)	4.9%	16.07 (5.88)	7.3%	11.10**
	Timeline cyclical	13.04 (3.39)	13.02 (3.40)	0.2%	12.82 (3.52)	1.7%	12.85 (3.31)	1.4%	0.39
	Illness coherence	13.77 (4.99)	12.72 (4.91)	7.6%	11.88 (4.61)	13.7%	11.41 (4.49)	17.1%	22.33**
	Treatment control	16.99 (3.36)	16.51 (3.52)	2.8%	16.58 (3.62)	2.4%	16.68 (3.55)	1.8%	9.13**
	Emotional representation	16.71 (5.23)	16.52 (5.20)	1.14%	15.69 (5.29)	6.1%	15.23 (5.18)	8.9%	18.24**
	Psychological attribution	11.93 (4.11)	***	***	11.94 (4.25)	0.1%	11.86 (3.94)	0.6%	1.03
	Risk factors	15.02 (4.15)	***	***	15.42 (3.95)	2.7%	15.53 (3.86)	3.4%	7.95**
	Immunity	5.33 (1.95)	***	***	5.29 (1.96)	0.8%	5.17 (1.86)	3.0%	0.08
	Accident/Chance	5.98 (1.92)	***	***	5.86 (1.92)	2.01%	6.00 (2.01)	0.3%	0.27

**** Not measured; **p<0.01; *p<0.05*

Between baseline and six-month follow-up, the most substantial mean change was again in disability (28.4%) but a large change was also seen in depression (22.2%), the illness perception of identity (what symptoms a patient experienced and whether they felt these were related to their back pain) (22.1%) and anxiety (20.1%). Much smaller changes were observed in the other variables, particularly the coping skills of diversion and cognitive coping, fear-avoidance beliefs, and the remaining illness perceptions.

Finally, between baseline and 12-month follow-up, the largest percentage mean change was again for disability (32.5%), depression (28.5%), identity (25.1%) and anxiety (23.6%), with the remaining variables showing only a very small mean change. The repeated measures ANOVAs showed that disability, self-efficacy, fear-avoidance beliefs, catastrophising, re-interpretation, anxiety, depression, acute-chronic timeline, identity, consequences, illness coherence, treatment control and emotional representation were all found to change significantly, while the remaining variables (cognitive coping, diversion, personal control, timeline cyclical, psychological attribution, immunity and accident/chance) did not.

This suggests that, even without manipulation through a specific treatment, there is potential for variables to change over time. Demonstration of change is an important pre-requisite for mediation. This change was not demonstrated in all the variables measured in the BeBACK study, and Table 5.3 also shows that the changes seen in this study, while statistically significant, were small in magnitude for several potential mediators. Diversion, cognitive coping, personal control, timeline cyclical, psychological attribution, accident/chance and immunity were not found to change substantially overall. This could indicate that they are stable characteristics that are not likely to change, but none were directly manipulated in this study. It could be that direct manipulation in an RCT would result in more change in all of the measured variables.

The next step in the analysis was to investigate whether change in the potential mediators was related to change in the outcome variable. Pearson's bivariate correlations were performed using the residualised change scores. The results showed that change in the self-efficacy measure had a fairly strong negative correlation with change in disability score (increasing self-efficacy was associated with reducing disability) and changes in catastrophising, anxiety, and depression had quite strong positive correlations (increasing, less favourable scores were associated with increasing disability), but that change in the remaining variables was only weakly associated with disability change at each follow-up point (see Table 5.4; variables that are shaded indicate a strong correlation). The criterion used to judge which factors were the strongest potential mediators was a combination of the correlations presented in Table 5.4 and the magnitude of change presented in Table 5.3, so that only variables which showed a large proportion of change as well as a strong correlation with change in outcome are highlighted.

Table 5.4 Associations between residualised changes in the potential mediator variables and residualised change in RMDQ score in the BeBACK study (Pearson's bivariate correlation coefficient)

Measure	Three-month follow-up	Six-month follow-up	12-month follow-up
Self-efficacy$^{\Delta}$	-.55**	-.52**	-.58**
Catastrophising$^{\Delta}$	.45**	.38**	.45**
Diversion$^{\Delta}$	***	.10*	.08
Re-Interpretation$^{\Delta}$	***	.03	.08
Cognitive Coping$^{\Delta}$	***	-.17**	-.18**
Fear-avoidance beliefs$^{\Delta}$	***	.31**	.33**
Anxiety$^{\Delta}$	***	.45**	.41**
Depression$^{\Delta}$	***	.57**	.55**
Personal control$^{\Delta}$	-.22**	-.25**	-.25**
Acute-chronic Timeline$^{\Delta}$	.14**	.18**	.21**
Identity$^{\Delta}$	***	.31**	.35**
Consequences$^{\Delta}$	.37**	.37**	.40**
Timeline Cyclical$^{\Delta}$	-.04	-.01	.01
Illness Coherence$^{\Delta}$	.12**	.09*	.06
Treatment Control$^{\Delta}$	-.14**	-.23**	-.20**
Emotional Representation$^{\Delta}$	.31**	.33*	.40**
Psychological Attribution$^{\Delta}$	***	.02	.13**
Risk factors$^{\Delta}$	***	.01	.04
Immunity$^{\Delta}$	***	.05	.10*
Accident/Chance$^{\Delta}$	***	.02	.09

$^{\Delta}$ *Residualised change score; *p<0.05; **p<0.01; *** Not available*

The above analysis suggests that while many of the tested variables change significantly over time, this change is not necessarily strongly associated with change in outcome. This is important, as if the change is not related to change in the outcome measure it is not likely to be a mediator of that change. Of all the variables tested, self-efficacy, catastrophising thoughts, anxiety and depression appear to be the best candidates for treatment mediation analysis.

5.8 Discussion

5.8.1 Summary

The BeBACK study was originally designed to test which psychological variables were predictive of future back pain outcomes such as disability and therefore of importance in estimating prognosis. The original findings indicated that a small number of variables were strong independent predictors of outcome (self-efficacy, personal control, acute-chronic timeline and identity), and when examining whether change in all the psychological variables was associated with change in disability, change in identity, self-efficacy and depression were found to be independently predictive. The secondary analysis presented here did not perform regression analysis as the aim was not to test the variables as predictors; instead, each variable was tested to investigate whether it was able to change over time and whether this change was associated with change in outcome, using residualised change scores rather than the raw change scores used by Foster et al (2010). The results of this secondary analysis suggest that catastrophising thoughts, identity, anxiety, and depression were the variables with the biggest mean score change over time and that residualised change in self-efficacy, catastrophising thoughts, anxiety, and depression are most strongly related to residualised change in disability. The discrepancy in results (catastrophising and anxiety were not considered to be potential mediators of disability by Foster et al) may be due to the type of change score used in the analysis. Foster et al (2010) used raw change scores while the results presented here used residualised change scores. The findings from the previous study may therefore be due to baseline score having a confounding effect. This is important to acknowledge and may add to the argument for residualised change scores being used more widely when conducting this type of analysis.

These findings show that although some variables can change over time, this change is not always strongly related to outcome and crucially, variables that have been shown to be strong predictors

are not necessarily candidates for mediation. This is particularly true of the illness perception variables which on average appeared to remain very stable over time. Clinically, this is an important finding as it suggests that interventions that focus only on trying to change variables that have been shown to be strong predictors may not show large treatment effects while interventions that attempt to change factors that have been shown to be modifiable even when not directly manipulated (such as in the observational data presented here) are likely to show much larger effects. It also highlights that the number of potential mediator variables is likely to be smaller than the number of predictor variables, so the number of variables that may be suitable targets for treatment may be smaller than previously thought. This suggests that simpler interventions that focus on one or two factors may perhaps be more efficient than complex interventions that attempt to address a large number of factors. It is important to note that the proportion mean change was calculated using raw change, rather than residualised change; the residualised change will be far smaller, so the results from the Pearson's bivariate correlations are reflective of these smaller changes.

5.8.2 Comparison with previous findings

The finding that self-efficacy is a potential mediator of outcome was unsurprising. The evidence from previous observational studies described in Section 5.1.1 noted self-efficacy as a variable often tested for mediation, with positive findings overall. However, the BeBACK study by far provides the largest sample in which to test self-efficacy compared to the previous literature and also has the advantage of being a prospective cohort rather than a cross-sectional study, which is crucial to assess change. This adds further weight to the argument for self-efficacy being a potential mediator. This analysis also adds to the evidence base for anxiety, depression and catastrophising thoughts in particular being potential mediators of functional outcomes.

Whilst the changes seen over time in self-efficacy in this study were small (less than 12% at any time point), residualised change in this variable had one of the strongest relationships with residualised change in disability. It must be acknowledged that the measure of self-efficacy used in the BeBACK study is very closely correlated with the measure of disability used, as the questions used in both scales are very similar. This could potentially be the reason for the strong correlations that are consistently found between self-efficacy and function. Investigation of this relationship is beyond the scope of this thesis, but further investigation of the psychometric properties of these tools and whether the measures have good content and construct validity is warranted.

5.8.3 Comparison with theoretical framework

The original aim of the BeBACK study was to test illness perceptions as assessed via the IPQ-R, based on the Self-Regulation model (Leventhal et al 2003), by investigating how well they predict clinical outcome. The original study found several illness perceptions to be strong predictors, in line with other research that used the IPQ-R (Kaptein et al 2010; Bijsterbosch et al 2009; Frostholm et al 2007) and looked at illness perceptions over longer periods of time (two years (Frostholm et al 2007) and six years (Kaptein et al 2010; Bijsterbosch et al 2009)). However, the present study provided little support for illness perceptions being potential mediators. This could again be due to the type of analysis, as none of the other studies used residualised change scores, meaning that baseline score could be responsible for the predictive power seen. Also, these studies aimed to assess the predictive value of illness perceptions rather than whether they can potentially explain causal associations, similar to the original BeBACK study, so were likely to draw different conclusions to those found in the present analysis. Predictive factors can be stable over time, while mediating factors need to be modifiable in order for them to be successfully targeted during treatment.

5.8.4 Potential limitations of the BeBACK secondary analysis

This analysis was conducted on a primary care LBP dataset which included a large number of variables to test as potential mediators. It may be suggested that focusing on only the factors found to be independently predictive of outcome by the original BeBACK study might have been more informative, but the point of this analysis was to add to the existing evidence base and identify which variables could potentially be mediators of treatment outcome. As this could have led to spurious associations being found between the variables, the magnitude of the changes and strength of associations were focused on, rather than statistical significance which was tested across all time-points. Also, few studies of mediators of treatment outcome exist (see Chapter 4), so using observational study data to get a sense of what might potentially mediate outcome was seen to be important.

It is also not clear from the analysis whether the variables did not change because they cannot change (i.e. are stable characteristics), or simply because they have not been directly manipulated in any way. This is a limitation of all mediation studies using observational data, and could explain the mixed results seen for tested variables in previous literature. It should also be highlighted that this analysis does not provide evidence for causal associations between the potential mediators and outcome; while the data used is longitudinal, temporality was not established, and adjustment for confounding was not carried out as part of this analysis. This highlights the importance of formal tests of mediation being carried out in RCT data, where causality can be more firmly investigated.

5.9 Conclusion

Psychological factors have the potential to mediate outcome and have been shown to do so in previous literature. As demonstrated in this secondary analysis of the BeBack data, variables found to be predictive of outcome do not necessarily also change over time or are strongly related to change in outcome. This highlights the importance of identifying modifiable variables and testing them appropriately as potential mediators in future randomised studies.

Conducting mediation analysis in observational studies cannot answer the important question of whether actually conducting an intervention which is designed to manipulate variables such as self-efficacy can lead to change in outcome. The next step therefore is to conduct mediation analysis on an intervention study that specifically targeted psychological variables during treatment to investigate whether the direct manipulation of psychological variables has an impact on patient outcome.

Chapter 6: Using RCT Data to Investigate Potential Mediators of Treatment Outcome

Part I: STarT Back trial

6.1 Introduction

The systematic review outlined in Chapter 4 located only a small number of articles of mediation analysis in musculoskeletal (MSK) pain intervention studies across a range of settings. Limitations were highlighted with the methods of mediation analysis employed, the psychometric properties of the tools used to measure the variables of interest, a lack of theoretical basis and no or limited measurement of variables over time. However, the review highlighted the emerging evidence for a number of psychological mediators of pain-associated disability outcomes such as pain catastrophising and pain self-efficacy. This supports the BeBACK secondary analysis (Chapter 5) which demonstrated that changes in several psychological factors were associated with changes in disability. The present chapter will explore how specifically targeting some of these psychological variables during an intervention might have an impact on outcome within the context of a randomised controlled trial (RCT), and thus enable a formal test as to whether these variables are mediators of the effect of treatment.

6.1.1 The Stratified Targeted Treatment approach: STarT Back Trial

The STarT Back trial was an RCT that compared stratified care with current best care in primary care patients consulting with low back pain (LBP) (Hill et al 2011). Specifically, the authors investigated whether targeting treatment according to levels of risk of a poor outcome (prognostic stratification) resulted in greater functional improvements than non-targeted care. Allocation to either the treatment or control group was on a 2:1 ratio basis in order to increase the number of patients receiving the intervention (Hay et al 2008), using block randomisation. If

patients were assigned to stratified care, treatment was matched to the patient's prognostic profiles using a brief prognostic index, the STarT Back tool (Hill et al 2008). This tool consists of nine questions relating to eight physical and psychological factors known to be predictors of persistent LBP-related disability (Table 6.1). Scores on both the overall tool score and the five-item psychological subscale score allocated patients to low-, medium- or high-risk targeted treatment groups (Table 6.2). Patients allocated to the low-risk group (low complexity) who received a score of three or fewer on the tool received a minimal package of evidence-based first-line care including advice and reassurance that their back problem was not serious and would resolve without intervention. Medium-risk patients, who received a score of four or higher on the tool but three or fewer on the psychological subscale were referred for standardised "best-practice" physiotherapy. A score of four or higher on the tool, and four or more on the psychological subscale specifically, meant that the patient was psychologically distressed and typically more complex to treat, and therefore at higher risk of a poor treatment outcome. These high-risk (complex) patients went on to receive psychologically informed physiotherapy (Main et al 2012). The high-risk intervention was delivered by physiotherapists who had undergone a series of training sessions (six days) focused on skills to help them address psychosocial barriers to patient recovery. The psychological factors discussed as part of this training included those measured in the STarT Back Tool such as fear-avoidance beliefs, catastrophising thoughts, anxiety and depression. Although the high-risk training was based on a cognitive-behavioural framework, the training did not constitute full-blown cognitive behaviour therapy (which would have required much more intensive training), but aimed to establish psychologically informed practice (PIP), where physiotherapists felt confident to address patients' unhelpful beliefs, emotions and behavioural responses to pain (Main et al 2012).

Table 6.1 STarT Back Tool items

Item (Where the question was taken from)	How Asked	How Measured
Referred leg pain (Single item)	My back pain has spread down my leg(s) at some time in the last 2 weeks	Disagree = 0; Agree = 1
Co-morbid pain (Single item)	I have had pain in the shoulder or neck at some time in the last 2 weeks	Disagree = 0; Agree = 1
Disability (Roland-Morris Disability Questionnaire (RMDQ))	I have only walked short distances because of my back pain	Disagree = 0; Agree = 1
	In the last 2 weeks, I have dressed more slowly because of my back pain	Disagree = 0; Agree = 1
Catastrophising thoughts (Pain Catastrophising Scale (PCS))	I feel that my back pain is terrible and it's never going to get any better	Disagree = 0; Agree = 1
Fear-avoidance beliefs (Tampa Scale for Kinesiophobia (TSK))	It's not really safe for a person with a condition like mine to be physically active	Disagree = 0; Agree = 1
Anxiety (Hospital Anxiety and Depression Scale (HADS))	Worrying thoughts have been going through my mind a lot of the time	Disagree = 0; Agree = 1
Depression (HADS)	In general I have not enjoyed all the things I used to enjoy	Disagree = 0; Agree = 1
Bothersomeness (Single item)	Overall, how bothersome has your back pain been in the last 2 weeks?	Likert scale 1-5 Not at all – Extremely (Dichotomised to 0 (0-3 on Likert scale) or 1 (4-5 on Likert scale))

Shaded items: Psychological subscale

Table 6.2 Scoring system for STarT Back tool

Score on STarT Back Tool	Risk Group Allocation
≥4/9 & ≥4 on Psychological subscale	High
≥4/9 & ≤3 on Psychological subscale	Medium
≤3/9	Low

At baseline, 851 patients were randomised (568 to intervention and 283 to control), with a mean age of 50 years (SD 14.8). Of this population 58.8% were female and 41.2% were male. Pain

duration was categorised in the same way as in the BeBACK dataset. In the present study similar proportions of patients were in each category at baseline (Less than one month =151 (17.7%), 1-3 months=190 (22.3%), 4-6 months=123 (14.5%), 7 months-3 years=209 (24.6%), More than 3 years=178 (20.9%)), suggesting similar numbers of patients would be classified as having acute or more persistent pain. For the high-risk group in particular (of interest in this thesis), the proportions were similar (Less than one month=22 (15.9%), 1-3 months=28 (20.3%), 4-6 months=23 (16.7%), 7 months-3 years=30 (21.7%), More than 3 years=35 (25.4%)). Descriptive data on the outcome measures and potential mediator variables is provided in Table 6.5 (page 162). Socio-demographic data were collected to assess baseline similarity across the intervention arms. The primary outcome in the STarT Back trial was back pain disability at 12 months (measured by the Roland-Morris Disability Questionnaire (RMDQ)), and secondary outcomes included catastrophising thoughts, fear-avoidance beliefs, anxiety and depression, health-related quality of life, general health (12-item Short-Form Health Survey (SF-12)), perception of change in back pain, pain intensity, number of pain-related days off work, satisfaction with treatment, and risk reduction using STarT Back subgroup (Hill et al 2011). In the study presented in this chapter, two measures of physical function (RMDQ and SF-12 – Physical Component Subscale (SF-12 PCS)) were selected as function is the outcome of interest in this thesis. The RMDQ is a measure related to back pain disability specifically, while the SF-12 PCS provides a more general measure of physical function. As the back pain-specific measure (RMDQ) relates more strongly to the outlined conceptual model presented in this chapter, only the results for this outcome will be presented here. The results for the SF-12 PCS are given in Appendix 6.1.

Participants allocated to the stratified intervention arm demonstrated a significant improvement in function at four- and 12-month follow-up compared to the control group receiving best current care. Of these two time-points the greatest differences in disability improvement were at four-

month follow-up (see Table 6.5, page 162). Table 6.5 presents the baseline scores and absolute change at four-month follow-up for both the intervention and control arms over time with the information split by prognostic subgroup.

6.1.2 Conceptual model

The high-risk training given to physiotherapists in the trial, although based on cognitive-behavioural principles, was not explicitly based on a theoretical model of pain or behaviour change. Although there is support for such models in whole or in part (see Chapter 1), it is likely that in a heterogeneous primary care population the level of individual patients' complexity will be variable. Even in pre-defined subgroups such as the high-risk group in this trial, where all patients scored highly on the brief psychological subscale of the screening tool, it is likely that patients vary widely in terms of the complexity and nature of their psychological distress. During the training, physiotherapists were taught to recognise psychological obstacles to recovery through their conversation with the patient and to deal with them accordingly (Main et al 2012) rather than explore and address each of the psychological factors in a systematic way using a formal measure. The intervention for high-risk patients was therefore hypothesised to work by giving therapists skills to reduce psychological distress through clear communication, reassurance and activity promotion, depending on how these factors were manifesting in an individual patient's story and causing the overall level of distress, in order to improve physical function. It therefore seemed appropriate to test whether the observed treatment effect was mediated by changes in patients' psychological distress (complexity) rather than by the individual psychological factors.

6.2 Aim

To test whether the wider concept of psychological distress (complexity) can be measured using the observed psychological variables measured in the STarT Back trial, and whether change in

psychological distress was a mediator of the beneficial stratified treatment effects for high-risk patients' disability outcomes observed at four-month follow-up.

6.3 Specific Objectives

1. To describe change over time (between baseline and four-month follow-up) for the functional outcome of interest (RMDQ);
2. To identify non-psychological variables that may potentially provide an alternative explanation for any mediation associations observed;
3. To assess the internal consistency of each of the psychological measures included in the mediation analysis;
4. To test each psychological distress variable targeted broadly by the intervention to ensure they meet the criteria for being a potential mediator;
5. To conduct mediation analysis to establish whether the latent variable of psychological distress is a mediator of disability outcome.

6.4 Hypothesis

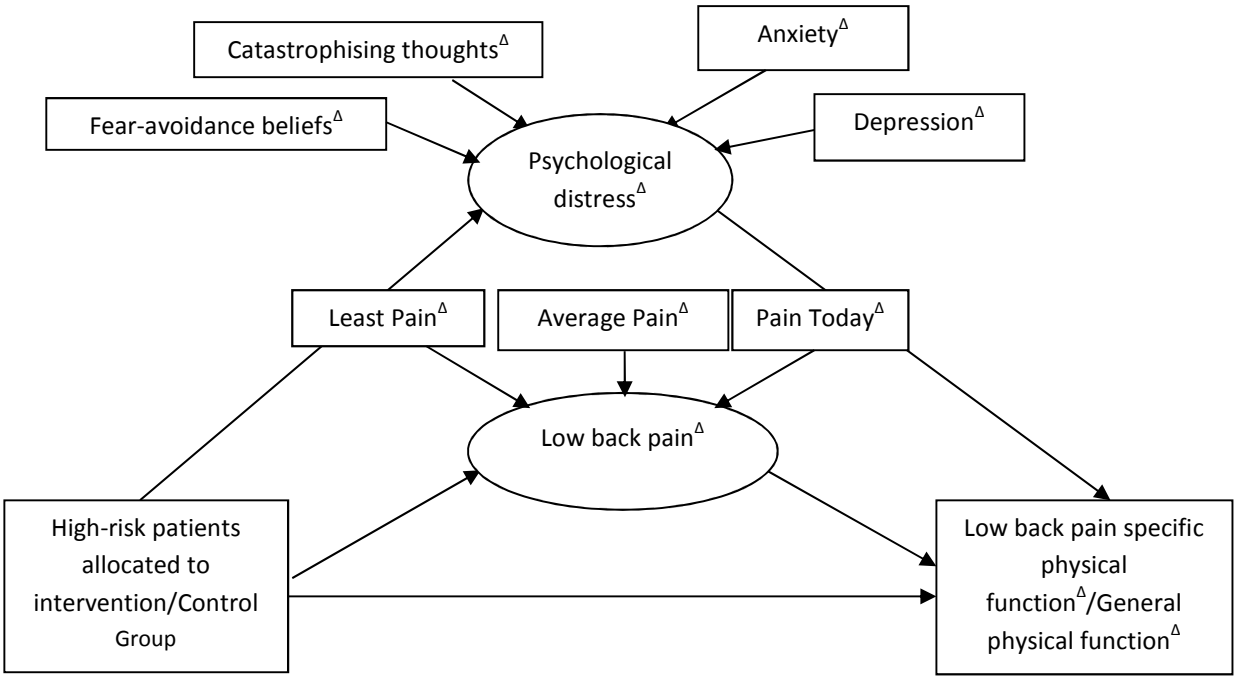
Improvements in disability scores of patients in the high-risk subgroup of the STarT Back trial were a consequence of change in psychological distress (complexity).

The high-risk group was focused on in this mediation analysis because patients in this subgroup received an intervention (psychologically informed physiotherapy) which specifically aimed to target psychological distress to reduce disability. The medium- and low-risk subgroups were also

subject to the same analysis, in order to allow for comparison across the whole trial population, although the results for these groups will only be briefly referred to in this chapter (see Appendix 6.2 for full details of the mediation models). It was hypothesised that any mediating effect found in these groups would be lower than that in the high-risk group, because psychological distress was not specifically targeted by treatment in the medium- and low-risk groups.

The final model to be tested is shown below (Figure 6.1). This analysis involves a number of steps which are outlined in Section 6.5, beginning with a description of structural equation modelling (SEM), the procedure used to carry out the analysis.

Figure 6.1 Final model to be tested



$^{\Delta}$*residualised change score*

6.5 Methodology

6.5.1 Structural Equation Modelling (SEM)

Much of the analysis in this chapter relates to structural equation modelling, or SEM. SEM is a confirmatory (theory based) approach that allows us to map out diagrammatically the paths between observed and unobserved variables, taking into account error variance (Byrne 2010). The benefits of using this technique for mediation analysis are set out in Chapter 3, but to summarise:

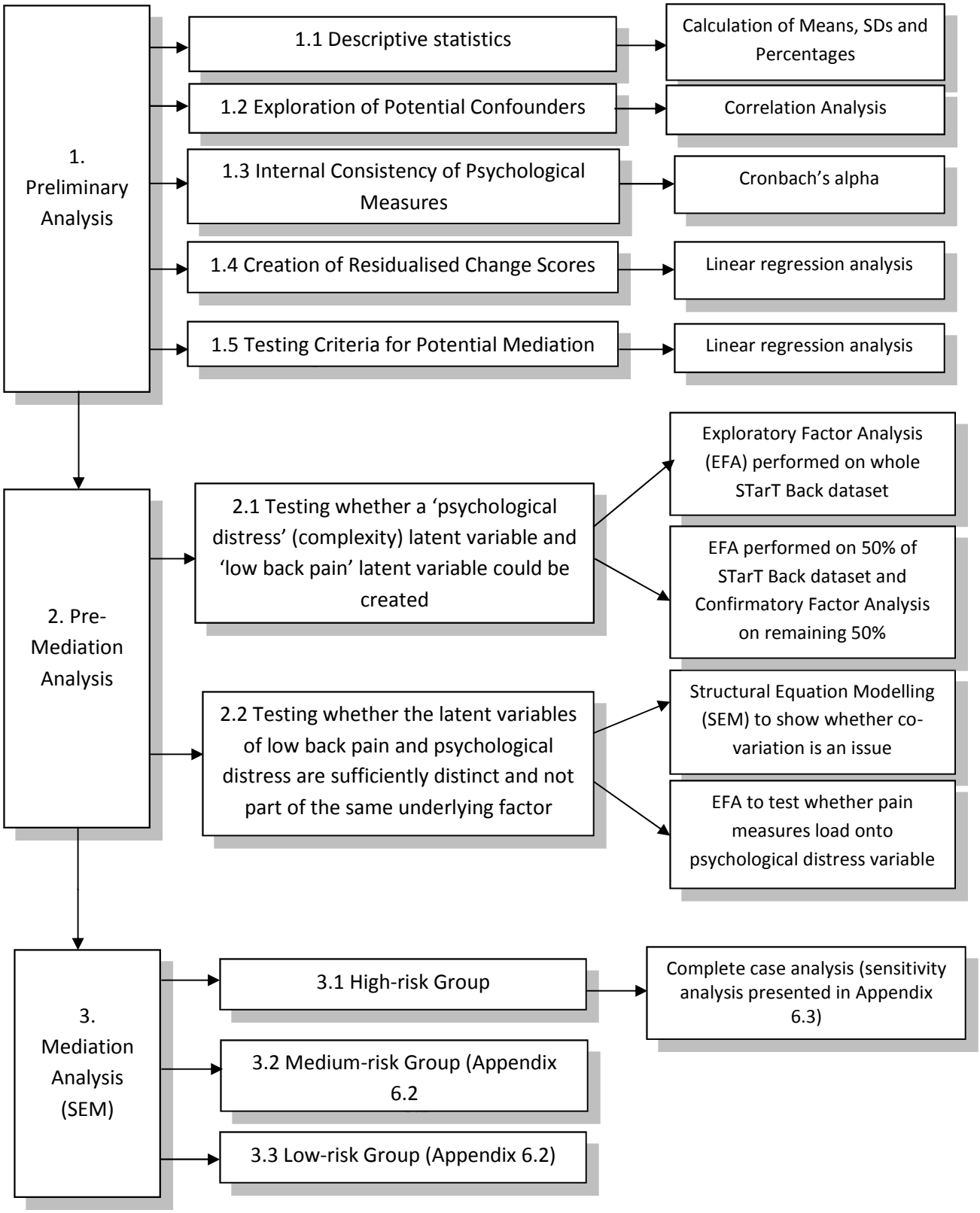
SEM combines linear regression and factor analysis (Tabachnick & Fidell 2007) which allows the inclusion of more than one variable as a measure of a particular construct, and helps account for measurement error; it can test more complex mediation models than traditional regression techniques; and it can compare different mediation models easily. SEM models comprise two parts: a measurement model, examining how the observed variables (psychological variables: in this analysis, anxiety, depression, catastrophising thoughts and fear-avoidance beliefs) relate to an unobserved (latent) psychological distress factor; and a structural model, examining the relationship between the unobserved factors only (Tomarken & Waller 2005). This latter part of the model separates SEM from confirmatory factor analysis (CFA), where only the measurement model is examined, but it is recommended that both parts of the model are analysed in a two-step approach (Anderson & Gerbing 1988; Cole & Maxwell 2003) so that each part can be assessed separately. In this chapter, SEM was chosen to perform the analysis as it allowed the testing of a latent distress variable (psychological distress) which is not possible when using standard regression techniques.

SEM can also provide the product of coefficients for the mediating effect with bias-corrected bootstrapped 95% confidence intervals (CIs), which is currently seen as an optimal way of

performing mediation analysis (see Chapter's 3 and 4). As explained in Chapter 3, bootstrapping is a way of re-sampling the data numerous times to create a new sampling distribution, which corrects for the often non-normal distribution of the indirect (mediating) effect. The number of bootstrapped samples recommended for performing this analysis is unclear, with papers reporting anything from 1,000 to 5,000 samples (e.g. Gaudiano et al 2010; Wicksell et al 2010). In this analysis, 1,000 bias-corrected bootstrapped samples were used. 95% CIs have been recommended over *p* values for assessment of the magnitude of the mediating effect (Hayes 2009). When performing mediation analysis using AMOS software, these CIs can only be calculated with complete case data. For this reason, only cases with complete data for all of the above variables were used in the analysis. This significantly reduced the number of cases included in the analysis (*n*=138 rather than the 236 patients who were in the high-risk group). Imputation was considered but it was not clear how latent variables (which were generated as part of the analysis) would be affected by imputation. As a sensitivity analysis, SEM model estimates without confidence intervals were generated from an analysis using the full sample (*n*=236 in the high-risk group), which are provided for comparison in Appendix 6.3.

All analyses were conducted using SPSS PASW statistics package version 18 and AMOS (add-on statistics package to SPSS) version 19. An overview of all the analyses performed can be seen in Figure 6.2.

Figure 6.2 Flowchart describing steps of analysis



6.5.2 Preliminary analysis (Step 1)

Prior to conducting the mediation analysis a number of preliminary analyses were required: descriptive statistics were presented for the variables to be used in the analysis (Step 1.1 of Figure 6.2); potential confounding factors were explored (1.2); the measurement properties of the included variables were examined (1.3); residualised change scores were created for use in the analysis (1.4); and investigations were carried out to test whether the psychological factor of interest met the basic requirements for mediation (1.5).

Descriptive statistics (means and standard deviations (SDs)) were calculated for baseline and four-month follow-up for the outcome and potential mediator variables for each of the three prognostic subgroups, and distributions checked for normality. The psychological factors emerging as the most promising potential mediators from the BeBack analysis in Chapter 5 were self-efficacy, catastrophising thoughts, anxiety and depression. Self-efficacy was not measured in STarT Back and therefore could not be included in this analysis. Fear-avoidance was considered to be an important prognostic factor and potential mediator by the STarT Back study team, and did meet the criteria for mediation in Chapter 5 although the associations were weaker than for the other factors mentioned above. Therefore, four of the psychological factors used in the STarT Back tool (catastrophising thoughts, fear-avoidance beliefs, anxiety and depression) were analysed via factor analysis to see whether they represented a single factor of psychological distress which could be used in the mediation model. Full scale reference standard measures of these constructs were used, rather than the single items used in the tool. For these four constructs, standardised measures were available that were deemed to have sound psychometric properties in LBP populations. Details of the outcome and potential mediator variables are presented in Table 6.3.

Table 6.3 Description of outcome and predictor measures

			Reference	How Scored
Outcome measure	Back specific physical function (Roland Morris Disability Questionnaire - RMDQ)		Roland & Morris (1983)	Add up number of 'Yes' rated responses – higher number of 'Yes' responses indicates higher level of disability
Potential mediators	Catastrophising thoughts (Pain Catastrophising Scale - PCS)		Sullivan et al (1995)	Higher score indicates high level of catastrophising
	Fear-avoidance beliefs Tampa Scale of Kinesiophobia - TSK)		Kori et al (1990)	4-point Likert scale 'strongly agree-strongly disagree' - Higher scores indicate higher fear-avoidance
	Anxiety (Hospital Anxiety and Depression Scale - HADS)		Zigmond & Snaith (1983)	Higher score indicates higher level of anxiety
	Depression (HADS)		Zigmond & Snaith (1983)	Higher score indicates higher level of depression
	Pain intensity	Least pain in last two weeks	-	11-point numeric rating scale (NRS) (0 = "no pain" to 10 = "pain as bad as could be")
		Average pain in last two weeks	-	11-point NRS (0 = "no pain" to 10 = "pain as bad as could be")
		Current pain intensity	-	11-point NRS (0 = "no pain" to 10 = "pain as bad as could be")
Potential confounders	Age		-	Actual age of patient recorded
	Sex		-	Coded as 1 (Female) or 2 (Male)
	Pain Duration ("How long since you have experienced a whole month without back pain?"		-	Coded as 1 (less than one month), 2 (one-three months), 3 (four-six months), 4 (seven months – three years) and 5 (more than three years)

Many factors are known to have an impact on function, and may provide an alternative explanation to any mediating effects of the psychological variables. Although STarT Back is an RCT and therefore patients in both groups should be similar at baseline, confounding can still occur in the *b* path of the mediation model (see Chapter 3). Age, sex (Thomas et al 1999) and pain duration (Bekkering et al 2005) have all been found to be associated with future outcomes, including disability, and may therefore act as confounders when analysing potential mediators (e.g. Smeets et al 2006; Kole-Snijders et al 1999; Morlion et al 2011). Although the evidence for their effect is mixed, with studies reporting inconsistent results regarding the predictive value of these variables, it was decided that they should be tested as possible confounders by analysing their relationship with each of the outcome measures (Step 1.2).

Pain intensity is thought to have an important role in patient outcomes (Morley et al 1999), is a strong prognostic factor, and often included as an outcome measure in studies of prediction and mediation. In the original STarT Back trial (Hill et al 2011), pain intensity was not specifically the focus of the high-risk training, but in primary care populations where many patients consult with acute pain, pain is a focus of treatment. Pain is strongly linked with psychological factors; the biopsychosocial model of pain described in Chapter 1 (Section 1.2) emphasised the multidimensional aspects of pain and that psychological factors are closely interlinked with the pain experience (e.g. Turk & Okifuji 2002). Pain intensity does therefore need to be examined as a potential alternative explanation to any mediating effects of the psychological variables. Pain intensity was measured during the STarT Back trial using three separate 11-point numerical rating scales which asked about least and average pain over the last two weeks, and pain on the day the participant completed the questionnaire. In this analysis the three separate scores of pain intensity available in the STarT Back dataset were explored using factor analysis to create a latent variable of low back pain.

Age, sex and pain duration were all previously identified as potentially being associated with outcome, and therefore might be confounders of any mediating effects of psychological distress. Pearson's bivariate correlations, point-biserial correlations and Spearman's correlations were used to examine the relationships between potential confounders and change in each of the outcome measures. Different types of correlation analyses were used depending on the different type of variable being examined; as age was measured as a numerical variable and was normally distributed (see Appendix 6.4) a Pearson's correlation was appropriate; as sex is a discrete dichotomy, a point-biserial correlation was carried out; and as pain duration was measured on a categorical scale, a Spearman's correlation, which is a non-parametric measure, was used to analyse this variable. Variables strongly related to residualised change in RMDQ scores were included in the mediation analysis as potential confounders.

Step 1.3 of the preliminary analysis involved examining the internal consistency of each of the psychological measures that were used to create the latent variable of psychological distress (complexity) using Cronbach's alpha (Cronbach 1951). Internal consistency is one way of assessing the unidimensionality of a measure, providing that the tool only represents a single scale and not multiple subscales (Field 2009). This is a potential issue with the TSK (fear-avoidance beliefs), the factor structure of which has been debated with 2- (Goubert et al 2004; Roelofs et al 2004; 2007) and 5-factor (Lundberg et al 2004) solutions being found previously. However, as this measure was used as a unidimensional construct in the original STarT Back trial, this is the structure that will be tested here. The two subscales of the HADS, representing anxiety and depression, were analysed separately. Testing internal consistency within this particular population is important as this measurement property is likely to vary somewhat between populations (Streiner 2003). A general guideline for interpreting the results of alpha is that a value of 0.70 or above is generally acceptable and suggests that the scale has appropriate levels of internal consistency within the

population being examined (Field 2009). This analysis was carried out using the whole scale or subscale of each measure at baseline in the STarT Back dataset.

Residualised change scores were then calculated for each potential mediator and outcome measure between baseline and four-month follow-up (Step 1.4). The rationale and procedure for this is described in Chapter 5 (Section 5.6.1). Residualised change scores are commonly used to look at mediating factors (e.g. George et al 2008; Skidmore et al 2015) but are more difficult to interpret. A sensitivity analysis using raw change scores is therefore provided for comparison with the results presented in this chapter (Appendix 6.5).

In order to check whether the four psychological factors met prerequisites for potential mediation, the absolute change that occurred between baseline and four-month follow-up was calculated and examined (Step 1.5). Linear regression analyses were then performed to investigate the relationships between residualised change in each of the identified potential mediator variables and residualised change in the outcome measure, and also to investigate the relationships between treatment group allocation (treatment or control) with change in each of the potential mediators. The three measures of pain intensity were also tested in this way to assess their relationships with the disability outcome measure, and investigate whether pain intensity could potentially mediate treatment effect.

6.5.3 Pre-mediation analysis (Step 2)

Structural Equation Modelling (SEM) was used to conduct the mediation analysis on this dataset. This type of analysis requires multiple variables or items per factor (latent variable), which allows the factor to be measured with greater reliability (Anderson & Gerbing 1988; Stephenson &

Holbert 2003). Exploratory factor analysis (EFA) and subsequently confirmatory factor analysis (CFA) was conducted with the four potential psychological mediators to see if the different measures represented a single 'psychological distress' factor (Step 2.1 of Figure 6.2). EFA was also conducted on the pain intensity measures to see if they represented a single factor of 'low back pain'. However, as EFA requires at least three measures of a potential latent factor, this analysis could not be repeated with the outcome measures (only RMDQ and SF-12 PCS were available as measures of physical function/disability). Each outcome was therefore investigated separately using observed rather than latent factors (RMDQ presented here; SF-12 PCS in Appendix 6.1).

The aim of exploratory factor analysis (EFA) is to identify the latent constructs that represent a set of measured variables (Fabrigar et al 1999). This is the difference between EFA and principal components analysis (PCA), for which the primary aim is to reduce the number of measured variables to a smaller number of components. In this study, since the objective is to explore whether potential mediators that were measured can be represented by a single factor, EFA is appropriate. The number of variables (or items) required for each factor is a matter of debate, with some authors suggesting six items (Fabrigar et al 1999) and others suggesting at least three (DeCoster 1998). The more items that are included, the more precise the EFA will be (Tinsley & Tinsley 1987).

Two important considerations when undertaking EFA are the number of cases available for analysis and the degree of correlation between the items or scales being analysed. There are a number of tests available to examine these; in this analysis, the Kaiser-Meyer-Olkin (KMO) statistic (Kaiser 1970, cited in Field 2009) was used to test whether the number of cases was adequate for the analysis to be performed, and Bartlett's test of sphericity tests whether the correlations were too low for the analysis to be valid (Field 2009). The KMO statistic produces a

value between 0 and 1, with values of 0.70 or above seen as showing an acceptable number of cases for analysis (Field 2009), while a statistically significant Bartlett's test indicates that correlations between the items are large enough for the factor analysis to be robust. However, Field (2009) acknowledges that the statistical significance of this test is dependent on sample size, so the correlation matrices should be checked by eye for any correlations being too small or very high. He also acknowledges the subjectivity of checking the size of correlations, and points out that although there are some guidelines relating to coefficients that are too small or too large, such values will be dependent on sample size and other properties unique to the dataset being analysed. This emphasises the importance of ensuring that the data used for EFA is sufficiently robust for this type of analysis.

There are several different procedures that can be used to fit an EFA model. Although all procedures will produce similar results, Maximum Likelihood (ML) provides the most information relating to inferential statistics such as model fit indices and confidence intervals (Fabrigar et al 1999) and is the recommended extraction method (Costello & Osborne 2005). ML has therefore been chosen for this particular analysis.

When an EFA is undertaken, several different tests can be used to determine the number of factors to extract (Fabrigar et al 1999), which include the estimation of eigenvalues (a measure of the explained variance of a factor (Kahn 2006)), where values above 1 are generally extracted, and Cattell's scree plot, in which the eigenvalues are plotted onto a graph in descending order. The point where the graph levels out is judged to be attributable to error, while any factors plotted before that point are judged to be factors (Tinsley & Tinsley 1987). Both these measures are widely used and accepted, and considering both tests together may provide a better solution than relying on one test (Tinsley & Tinsley 1987). A further test which may provide a more

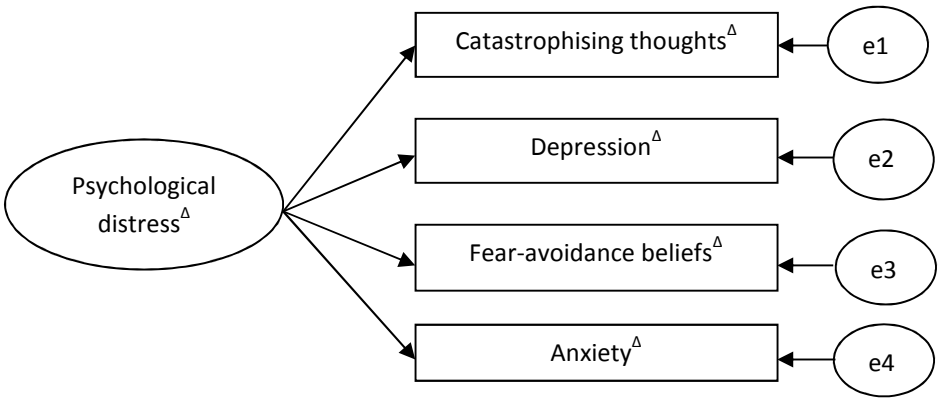
accurate result is parallel analysis (Velicer & Jackson 1990, cited in Costello & Osborne 2005; Kahn 2006), in which a combination of both eigenvalues and the scree plot from a random dataset with the same number of items and cases are plotted (Russell 2002) and compared with those generated from the real data (Schmitt 2011). Parallel analysis is recommended as the best method to use (Schmitt 2011), however it is for use on item-level data which were not used in this analysis (see below for further detail). In the present analysis therefore, the eigenvalues and scree plot were used to judge the results of the analysis.

Finally, factor solutions must often be rotated to try and make sense of the factors identified. There are a number of different methods of factor rotation available, depending on the data being analysed. Orthogonal rotation methods assume that any generated factors are uncorrelated, while oblique methods allow correlated factors to be used (Fabrigar et al 1999). Oblique methods of rotation are likely to provide more realistic solutions and estimate correlations among factors, which can aid interpretation (Fabrigar et al 1999) and are most often recommended for use with psychological and social science data (Schmitt 2011; Gaskin & Happell 2014). In fact some have argued that orthogonal methods should never be used when analysing psychological data (Field 2009). When using large data sets, the values computed using the various methods are almost identical and so comparisons between results of studies using different rotational methods are only marginally affected by the type of rotation. There are several methods of oblique rotation available, which have been reported to produce similar results (Fabrigar et al 1999). In this analysis, oblique rotation using direct oblimin was preferred, as this has been reported to be an appropriate rotation method for psychological variables showing correlation (Conway & Huffcut 2003).

EFA was firstly conducted on the STarT Back dataset as a whole, and then repeated using 50% of cases that were randomly selected from the STarT Back dataset. Residualised change scores for catastrophising thoughts (PCS), fear-avoidance beliefs (TSK) and anxiety and depression (HADS) were entered. The residualised change scores were used instead of full baseline measures because the full SEM models were used to test whether *change* in psychological distress was related to change in outcome. The final analysis involved a CFA which was performed on the remaining 50% of the dataset not used in the EFA. This analysis was repeated to create a latent variable for low back pain, using residualised change in the three measures of pain intensity used in the STarT Back trial (least pain in last two weeks, average pain in last two weeks and current pain).

CFA is used to test a hypothesised model between observed and latent variables (Byrne 2010). CFA forms the measurement model of SEM. Unlike in EFA, the number of factors is pre-determined based on theory and/or previous research. EFA and CFA are often used together, with EFA being conducted on one dataset and CFA being used on another, in order to verify the structure created in EFA (Fabrigar et al 1999). Alternatively, a dataset can be randomly split to allow an EFA to be performed on one half, and a CFA to be performed on the other. The statistical package AMOS allows a hypothesised model such as that shown in Figure 6.3 to be tested. The hypothesised latent variable is represented by a circle, the measured variables are represented by rectangles, and error is also accounted for in the model (represented by circles). Unidirectional arrows represent the *hypothesised* effect of one variable on another, as causality cannot be determined using this type of analysis.

Figure 6.3 Example of CFA modelling in AMOS



In SEM, the software will only test paths where they are specifically added to the model. If no path between particular variables is present, then it assumes there is no relationship between the two variables (Schumacker & Lomax 2010). If there was in fact a relationship, then this may lead to a poorly-fitting model, so relationships between variables of interest must be tested to ensure they are correctly specified to have (or not have) a relationship in the model. A strong correlation may also mean that the two latent variables are representing the same factor, and would be better represented as a single, overall latent variable. It was therefore important to test whether the two latent variables of psychological distress and low back pain were correlated (Step 2b), as a strong correlation (i.e. 0.60 or above) would need to be acknowledged in the model by means of a double-headed arrow to show co-variance.

To test for co-variance between psychological distress and low back pain, the latent variables were analysed using CFA to generate a correlation between them. If a strong relationship was found, an EFA would be performed to see whether the observed pain intensity measures also load onto and are represented by the same psychological distress factor.

6.5.4 Mediation analysis (Step 3)

SEM, as outlined in Section 6.5.1, allows the relationships between observed and latent variables to be analysed. The CFA carried out as part of the pre-mediation analysis makes up the measurement model in SEM. Full SEM involves testing the relationships between the measurement model(s) and other variables of interest, which together make up the structural model (Step 3.1).

The output from SEM provides several goodness-of-fit indices that can be used to assess how well the proposed model fits the available data. The most common of these is the Chi-square (X^2 (CMIN)) statistic, which tests whether the proposed model exactly fits the data tested (Tomarken & Waller 2003). However, this test is seen as overly strict in testing for exact fit and it is also affected by sample size, with large samples often resulting in 'bad' fit (Tomarken & Waller 2003). It also gives less information about how well the model fits the data compared with other indices (Holbert & Stephenson 2002; Tomarken & Waller 2003). Dividing the X^2 value by the degrees of freedom in the model (CMIN/df) is suggested as a way to reduce the dependence of the X^2 on sample size (Truchon et al 2008). A number of alternative indices have also been proposed, including the goodness of fit index (GFI), normed fit index (NFI), comparative fit index (CFI), root mean squared error of approximation (RMSEA) and standardised root mean squared residual (SRMR). The GFI and NFI have been dismissed by some as poor indices (Hu & Bentler 1998; Hu & Bentler 1999) because of issues related to sample size and insensitivity (MacCullum & Austin 2000). A number of different indices are necessary to report as each has different strengths and weaknesses, and the use of several complementary statistics is recommended (Bentler 2007; Byrne 2010). For example, SRMR and RMSEA are recommended with samples over 250 and SRMR in combination with any of the other indices are recommended for use with smaller samples

(Holbert & Stephenson 2002). The RMSEA in particular gives more information about model fit by providing confidence intervals around the value. It also provides a *p* value for the model's closeness of fit (PCLOSE).The suggested cut-offs for each of the indices that will be used in the present analyses are given in Table 6.4 below.

Table 6.4 Goodness-of-fit statistics for SEM and their cut-off values

Index	**Description**	**Additional Values to Include**	**Indication of Good Model Fit**
X^2 (Chi Square)	The likelihood of the specified model being true	X^2/df (CMIN)	$X^2 = p<0.05$ X^2/df = between 2-5
CFI (Comparative Fit Index)	Comparison of X^2 values of specified and null models	None	>0.95 = adequate fit
RMSEA (Root Mean Square Error of Approximation)	Measure of the degree of misfit in the proposed model (Chen et al 2008)	PCLOSE, CIs	<0.05 = good fit <0.08 = acceptable fit (Byrne et al 2008)
SRMR (Standardised Root Mean Square Residual)	Absolute measure of fit – standardised difference between the observed and predicted correlations	None	<0.08 = acceptable fit

Several authors have advised caution when examining the fit statistics, as they have no empirical basis (McDonald & Ho 2002), with some authors going as far as to say their use is meaningless (Barrett 2007). There is no way of choosing which fit statistic is 'superior' and therefore it is difficult to make a decision if one index suggests good fit and others do not (McDonald & Ho 2002). Others have noted that goodness-of-fit indices may relate to the measurement model rather than the structural model, meaning that the indices might be misrepresentative (O'Boyle & Williams 2011). Therefore, these indices must be interpreted with caution and seen as a guide to model fit, rather than giving an objective conclusion.

SEMs were generated for each of the risk strata in the STarT Back trial (low, medium and high) with both treatment and control cases being included within the model. The outcomes of back pain-specific and general physical function were each represented by a single measure, as no other variables measuring these constructs were available in the dataset. It was therefore assumed that the observed variable provided a perfect measure of the construct (Schumacker & Lomax 2010). Models are presented for each outcome separately (SF-12 PCS in Appendix 6.1).

6.6 Results

6.6.1 Preliminary analysis

Table 6.5 contains the mean scores at baseline and mean change at four-month follow-up for all of the potential mediator variables. There was a difference between the amounts of change in potential mediators observed in the treatment groups at four-month follow-up compared to the control group. These differences in change were most marked in the high-risk group, where scores were higher at baseline, with a 4.4 point difference in change being observed between treatment and control for catastrophising thoughts and a 5.8 point difference in change for fear-avoidance beliefs. Anxiety, depression and the three pain measures showed less substantial differences. The SDs for all mediators were quite large, suggesting large variability in change within the high-risk group.

As all further analysis is based on parametric tests that assume that the data are normally distributed, it is important that this assumption is tested. Kurtosis and skewness values were therefore examined, along with histograms and normal probability plots (see Appendix 6.6). Analyses were conducted on baseline and four-month follow-up scores, but also on the residualised change scores as these were used in the majority of the analysis. For the treatment

arm in the high-risk group, skewness values were between 0.02 and -0.95, and kurtosis values were between -0.08 and 1.46. In the control group, skewness values ranged from -1.18 to 1.02, and kurtosis values ranged from -1.11 to 1.84. These values suggest that there is some deviation from normality in some of the variables to be used for analysis. Inspection of the histograms and probability plots confirmed the non-normality of some of the variables, although the probability plots do not show such a strong departure. These tests indicate that the findings presented here may not be generalisable to other populations as the estimates produced in the regression and SEM models that follow in this chapter may not be robust (Field 2009).

Table 6.5 Baseline means and SDs and mean change at four-month follow-up for potential mediator variables in the STarT Back dataset

		Baseline Score (Mean and SD)						**Four-month follow-up (Mean change and SD)**					
		Treatment (*n*=370)			**Control (*n*=179)**			**Treatment (*n*=370)**			**Control (*n*=179)**		
		L (*n*=100)	**M (*n*=177)**	**H (*n*=93)**	**L (*n*=50)**	**M (*n*=84)**	**H (*n*=45)**	**L (*n*=100)**	**M (*n*=177)**	**H (*n*=93)**	**L (*n*=50)**	**M (*n*=84)**	**H (*n*=45)**
Outcome													
Disability		4.08 (3.20)	9.93 (4.65)	14.41 (4.31)	3.96 (3.60)	9.75 (4.66)	14.07 (4.88)	1.52 (3.21)	5.23 (5.67)	7.49 (6.48)	0.80 (4.04)	3.98 (4.80)	3.62 (4.38)
Potential Mediators													
Catastrophising thoughts		7.89 (5.31)	14.79 (8.00)	25.24 (10.49)	7.28 (5.48)	12.82 (8.22)	25.88 (10.54)	2.83 (5.93)	6.09 (8.05)	11.07 (12.95)	0.84 (8.01)	4.37 (8.42)	6.64 (10.26)
Fear-avoidance beliefs		36.07 (4.70)	39.25 (4.63)	46.21 (5.17)	35.96 (5.98)	39.72 (4.58)	45.52 (5.85)	3.47 (5.63)	5.07 (6.15)	9.24 (7.56)	3.00 (6.23)	3.22 (4.13)	3.40 (4.68)
Anxiety		4.93 (2.93)	6.85 (3.58)	10.01 (4.39)	5.56 (3.63)	7.10 (3.77)	10.31 (3.59)	0.61 (2.50)	1.77 (3.20)	3.39 (4.10)	1.28 (3.03)	0.72 (3.09)	2.49 (3.95)
Depression		3.06 (2.71)	5.44 (3.19)	8.77 (4.34)	2.86 (2.57)	5.60 (3.84)	8.40 (3.70)	0.52 (2.19)	1.81 (3.32)	3.55 (4.05)	0.14 (2.75)	1.16 (3.31)	1.69 (3.55)
Pain Intensity	Least	2.44 (1.83)	4.25 (2.48)	6.16 (2.58)	2.46 (1.85)	4.04 (2.61)	5.96 (3.25)	0.76 (2.01)	1.89 (2.88)	2.98 (2.87)	0.48 (2.10)	1.56 (2.49)	1.76 (3.19)
	Average	4.45 (2.08)	6.99 (2.05)	7.72 (2.12)	4.62 (2.36)	6.82 (2.06)	8.18 (1.80)	1.64 (2.81)	3.36 (3.15)	3.90 (3.24)	1.50 (2.38)	2.82 (3.09)	2.56 (2.62)
	Current	2.98 (2.05)	4.95 (2.41)	6.40 (2.33)	2.98 (2.10)	4.48 (2.43)	6.51 (2.64)	1.30 (2.25)	2.40 (2.95)	3.00 (2.88)	0.80 (2.45)	1.44 (2.75)	1.62 (3.00)

L=Low-risk; M-Medium-risk; H=High-risk

The next stage of the analysis was to investigate whether age, gender, and pain duration could be potential confounders of the results of the mediation analysis. Pearson's correlations were conducted with baseline values of age, point bi-serial correlations with sex and Spearman's correlations with pain duration and residualised change in low back pain disability (RMDQ) and general physical function (SF-12 PCS). The results in Table 6.6 show that age and sex had very weak relationships with each outcome measure in that the correlation coefficients were below 0.50 and not statistically significant for either treatment arm or any risk group. For pain duration, the coefficients were generally higher (longer pain duration was associated with increased disability), especially in the high-risk group, and some of these coefficients were statistically significant, but were small in magnitude. Given these findings it was decided not to include these potential confounders in the subsequent mediation analysis.

Table 6.6 Correlation coefficients for the association of potential confounding variables with change in RMDQ and SF-12 PCS

Outcome	**Variable**	**Treatment Allocation**	**Four-month follow-up**		
			Low-risk group	**Medium-risk group**	**High-risk group**
RMDQ$^{\Delta}$	Age	Treatment	-0.01	0.11	0.08
	Sex		0.11	-0.09	-0.02
	Pain duration		0.06	0.19	0.33**
	Age	Control	-0.05	0.01	-0.00
	Sex		-0.00	-0.17	0.22
	Pain duration		0.04	0.28*	0.37*

*p<0.05; **p<0.01, $^{\Delta}$residualised change*

The test of Cronbach's alpha for each included measure, carried out on the entire dataset rather than just the high-risk group, showed that each measure met the criteria for internal consistency (Cronbach's alpha of 0.70 or above), suggesting that they are appropriate for use in the SEM

analysis. Table 6.7 below gives the alpha value for each measure. These values are similar to those calculated in other LBP populations (e.g. 0.84 for the 17-item TSK (French et al 2007 – secondary care population); 0.70 for the Dutch version of the 17-item TSK (Swinkels-Meewisse et al 2003 – primary care population); 0.92 for the PCS (Osman et al 2000 – chronic pain outpatient population)).

Table 6.7 Cronbach's alpha values for each psychological measure

Measure	**Number of cases analysis based on**	**Cronbach's alpha value**
Pain Catastrophising Scale (PCS)	839	0.94
Tampa Scale of Kinesiophobia (TSK)	819	0.73
Hospital Anxiety and Depression Scale – Depression (HADS-D)	851	0.85
Hospital Anxiety and Depression Scale – Anxiety (HADS-A)	850	0.82

After calculating residualised change scores, linear regression analyses for all of the potential mediator variables with both of the outcome measures were performed (Table 6.8; results for the SF-12 PCS presented in Appendix 6.1). The results indicated that in the high-risk treatment group, residualised change in the psychological variables strongly predicted residualised change in RMDQ and accounted for between 25% and 39% of the variance in this outcome. However, residualised change in the pain variables included for comparison with the psychological variables were shown to be stronger predictors; they accounted for between 51% and 63% of the variance of residualised change in RMDQ. In the control group, all potential mediators also accounted for a large amount of variance of residualised change in RMDQ. This shows support for the psychological factors as well as the pain variables to potentially mediate treatment outcome, as these variables do show change over time and are associated with outcome (*b* path), especially in the high-risk subgroup.

Table 6.8 Univariable associations of changes in each potential mediator with change in functional outcomes in STarT Back participants: Linear regression analyses – High-risk group

Outcome	Predictor	Treatment Allocation	Change at four-month follow-up			
			Unstandardised B (SE)	95% CI	Standardised β	R-square change
RMDQ$^{\Delta}$	Catastrophising thoughts$^{\Delta}$	Treatment (*n*=93)	0.49 (0.09)	0.31 to 0.67	0.50	0.25
	Fear-avoidance beliefs$^{\Delta}$		0.57 (0.09)	0.40 to 0.74	0.57	0.33
	Anxiety$^{\Delta}$		0.59 (0.09)	0.40 to 0.77	0.56	0.31
	Depression$^{\Delta}$		0.66 (0.09)	0.48 to 0.83	0.62	0.39
	Least pain$^{\Delta}$		0.77 (0.08)	0.62 to 0.93	0.72	0.51
	Average pain$^{\Delta}$		0.77 (0.07)	0.63 to 0.91	0.74	0.55
	Current pain$^{\Delta}$		0.84 (0.07)	0.71 to 0.98	0.80	0.63
	Catastrophising thoughts$^{\Delta}$	Control (*n*=45)	0.58 (0.11)	0.37 to 0.80	0.64	0.41
	Fear-avoidance beliefs$^{\Delta}$		0.73 (0.12)	0.05 to 0.97	0.67	0.45
	Anxiety$^{\Delta}$		0.46 (0.10)	0.27 to 0.65	0.59	0.35
	Depression$^{\Delta}$		0.58 (0.09)	0.40 to 0.75	0.71	0.50
	Least pain$^{\Delta}$		0.45 (0.10)	0.26 to 0.65	0.59	0.34
	Average pain$^{\Delta}$		0.50 (0.11)	0.28 to 0.72	0.57	0.32
	Current pain$^{\Delta}$		0.47 (0.10)	0.27 to 0.68	0.58	0.34

$^{\Delta}$*Residualised change scores*

Finally, it was important to also test the *a* path, or the relationship between treatment allocation and each of the potential mediators. If the treatment had little effect on the potential mediating variables then the variables are unlikely to be the mechanism through which the treatment was successful. Linear regression analyses were performed with treatment allocation as the predictor variable and residualised change in each of the potential mediators as outcome variables. The results in Table 6.9 show that a small proportion of variance (between 2 and 12%) of residualised change is explained by treatment allocation (stratified care versus control) for each of the potential mediators. For residualised change in anxiety in particular, the beta value is very small and the association is not statistically significant, as indicated by the 95% CI. The standardised beta values indicate that treatment allocation has the strongest association with residualised change in fear-avoidance beliefs.

Table 6.9 Univariable associations of change in treatment allocation with change in each potential mediator: Linear regression analyses - High-risk group

Outcome	Predictor	Unstandardised B (SE)	95% CI	Standardised β	R-square change
Catastrophising thoughts$^{\Delta}$	Treatment Allocation	0.41 (0.18)	0.06 to 0.76	0.19	0.04
Fear-avoidance beliefs$^{\Delta}$	Treatment Allocation	0.72 (0.17)	0.39 to 1.06	0.34	0.12
Anxiety$^{\Delta}$	Treatment Allocation	0.28 (0.18)	-0.08 to 0.63	0.13	0.02
Depression$^{\Delta}$	Treatment Allocation	0.47 (0.18)	0.12 to 0.82	0.22	0.05
Least pain$^{\Delta}$	Treatment Allocation	0.44 (0.18)	0.08 to 0.79	0.21	0.04
Average pain$^{\Delta}$	Treatment Allocation	0.58 (0.18)	0.23 to 0.92	0.27	0.07
Current pain$^{\Delta}$	Treatment Allocation	0.52 (0.18)	0.17 to 0.87	0.25	0.06

$^{\Delta}$*=residualised change*

In summary, these preliminary analyses show support for the hypothetical model, in that the four psychological variables to be included in the mediation model (catastrophising thoughts, fear-avoidance beliefs, anxiety and depression) changed significantly between baseline and follow-up and that residualised change in each of these measures is predictive of residualised change in disability in those receiving the intervention. Treatment allocation was also found to be predictive of residualised change in all of the potential mediator variables, with the exception of anxiety. All of these measures were also found to be internally consistent for the STarT Back population. Baseline measures of non-psychological variables (age, sex and pain duration) were found not to be strongly related to residualised change in either functional outcome measure, suggesting that they were unlikely to confound the results of the mediation analysis. These variables were therefore excluded from subsequent analyses. However, the measures of pain intensity did change over time and residualised change in each pain measure was predictive of residualised change in both outcome measures, indicating that these measures may potentially help explain

change in outcome. These variables were therefore taken forward as planned as an alternative mediating pathway of the relationship between treatment allocation and residualised change in functional outcome.

6.6.2 Pre-mediation analysis: Exploratory factor analysis of latent mediator variables

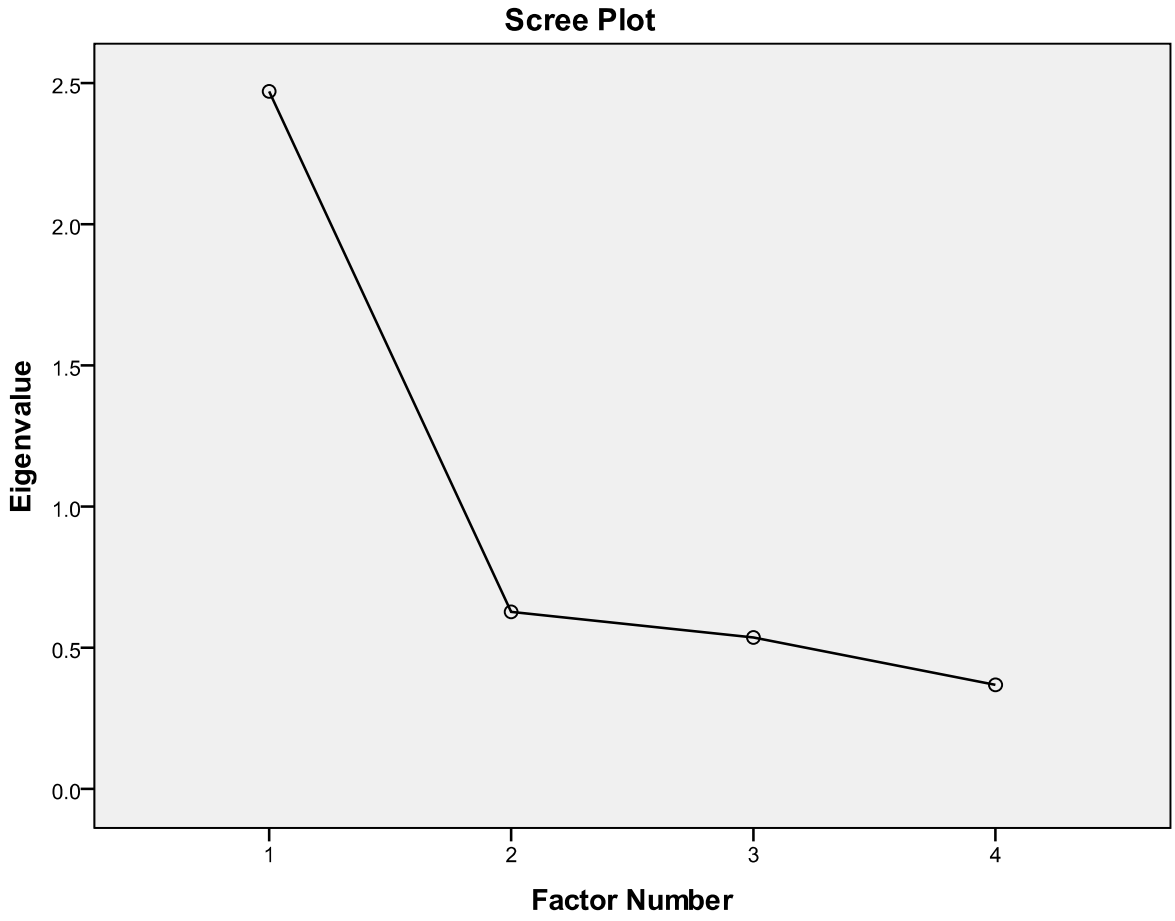
591 cases were included when the whole dataset was analysed, which is adequate for EFA. This was verified by the KMO statistic, which gave a value of 0.766. Bartlett's test of sphericity indicated that the correlations between the scores were also high enough for EFA to be performed ($X^2_{(6)} = 698.10$, $p<0.05$) (see Section 6.5.4 for an explanation of these tests).

For the psychological measures, one factor was extracted, confirmed by the scree plot (Figure 6.4), and accounted for 50% of the variance. The factor matrix shown below (Table 6.10a) indicates that all measures included in the analysis have strong factor loadings of well above 0.30 (a usual minimum requirement for retention (Field 2009)) with the extracted factor.

Table 6.10a Factor matrix (EFA) for the psychological distress latent variable: Entire STarT Back population (*n*=591)

Measure	**Factor loading**
Residualised change in depression (HADS-D)	0.82
Residualised change in anxiety (HADS-A)	0.73
Residualised change in catastrophising thoughts (PCS)	0.63
Residualised change in fear-avoidance beliefs (TSK)	0.62

Figure 6.4 Scree plot (EFA) for the psychological distress latent variable: Entire STarT Back population

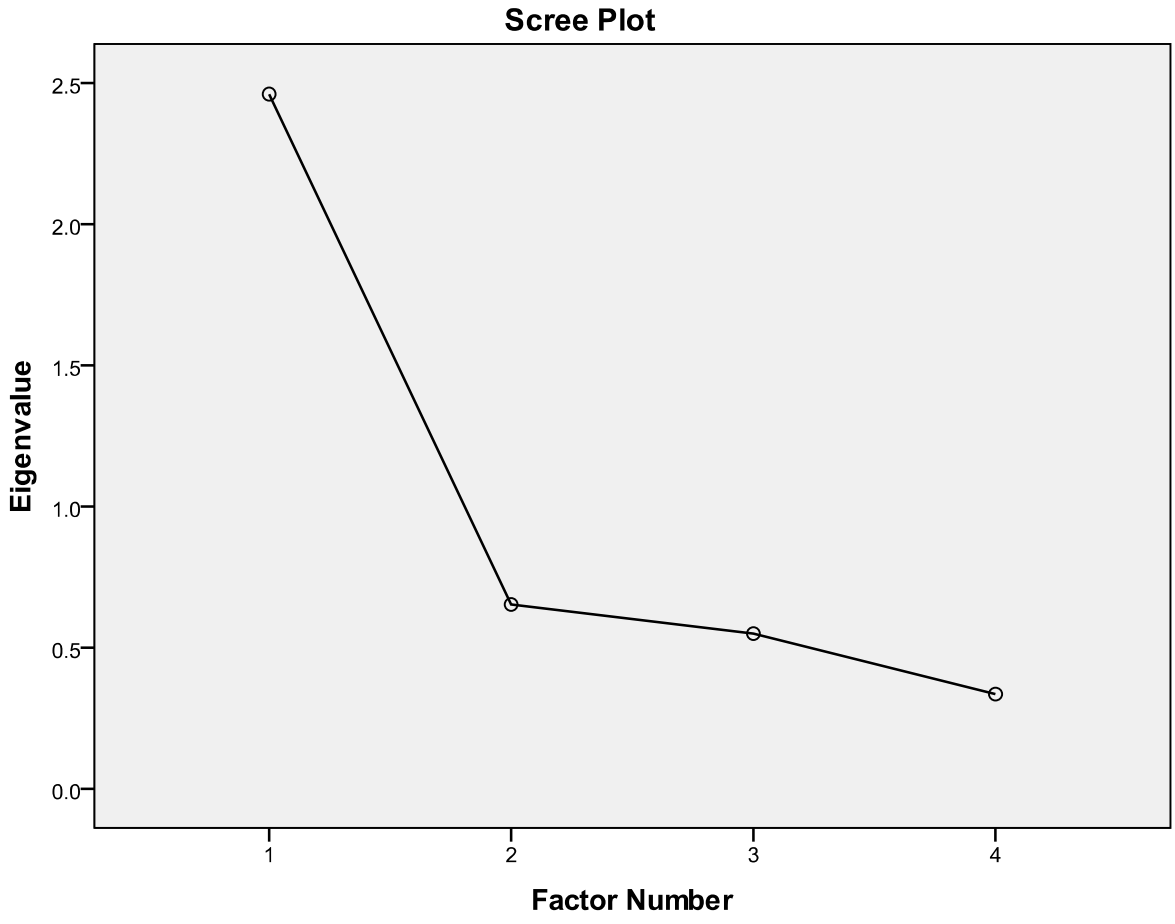


In order to perform a CFA of the generated models, the STarT Back dataset was split into two randomly selected halves using the 'Select Cases' function in SPSS. The software package selected 394 cases for the EFA. The KMO statistic of 0.753 suggested that this was adequate for EFA to be performed. Bartlett's test also indicated that correlations between variables were high ($X^2_{(6)} = 474.57$, $p<0.05$). Extracting one factor from this half of the dataset explained 50% of the variance. This was confirmed by the scree plot (Figure 6.5), and the factor loadings were similar to those generated for the whole dataset (Table 6.10b).

Table 6.10b Factor matrix (EFA) for the psychological distress latent variable: Randomly selected 50% of STarT Back population (*n*=394)

Measure	**Factor loading**
Residualised change in depression (HADS-D)	0.85
Residualised change in anxiety (HADS-A)	0.75
Residualised change in fear-avoidance beliefs (TSK)	0.61
Residualised change in catastrophising thoughts (PCS)	0.58

Figure 6.5 Scree plot (EFA) for the psychological distress latent variable: Randomly selected 50% of the STarT Back population

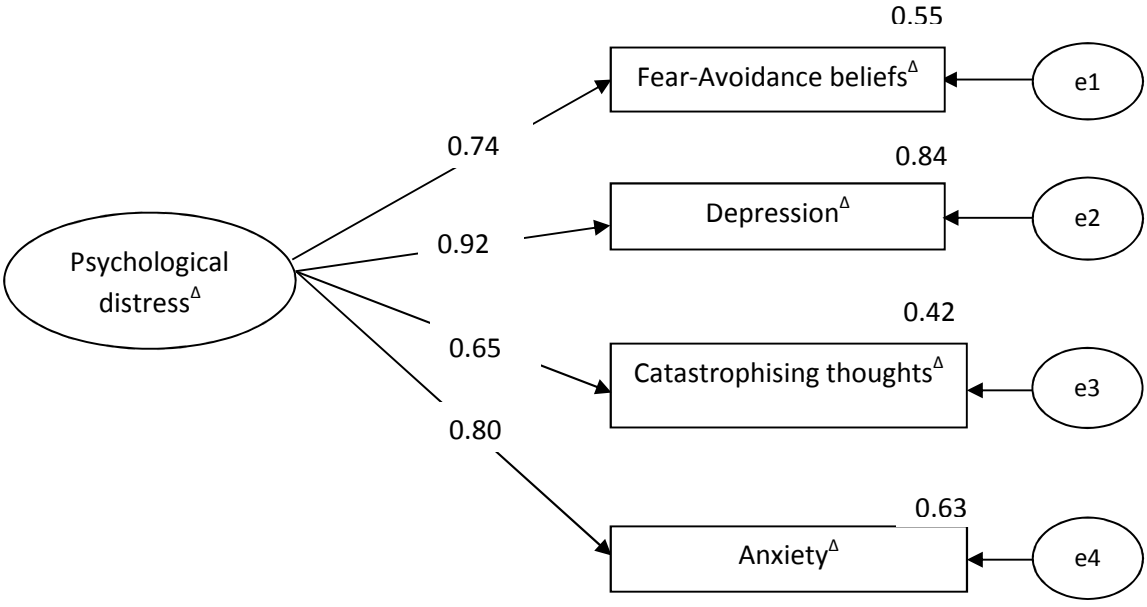


For the CFA, a model was generated with a single factor, labelled "Psychological distress" (Figure 6.6). The model fit statistics for the CFA are given below, all of which matched the criteria for adequate to good model fit with the exception of the CMIN/DF, so it was felt that overall the model provided an adequate fit to the data (Table 6.11). The results of the EFAs in both the full and random half of the dataset, as well as the CFA suggest that it was appropriate to use a single psychological distress factor to represent the four potential psychological mediators.

Table 6.11 Model fit statistics for CFA of the psychological distress factor

Model Index	**Current Model (CIs)**	**Good Model Fit**
CMIN*	1.14	Non-significant result
DF	2	
P	.57	
CMIN/DF	0.57	Between 2-5
CFI	1.0	>0.95
RMSEA	0.00 (0.00-0.14) (PCLOSE 0.67)	<0.08
SRMR	0.01	<0.08

Figure 6.6 CFA of psychological distress as a latent variable

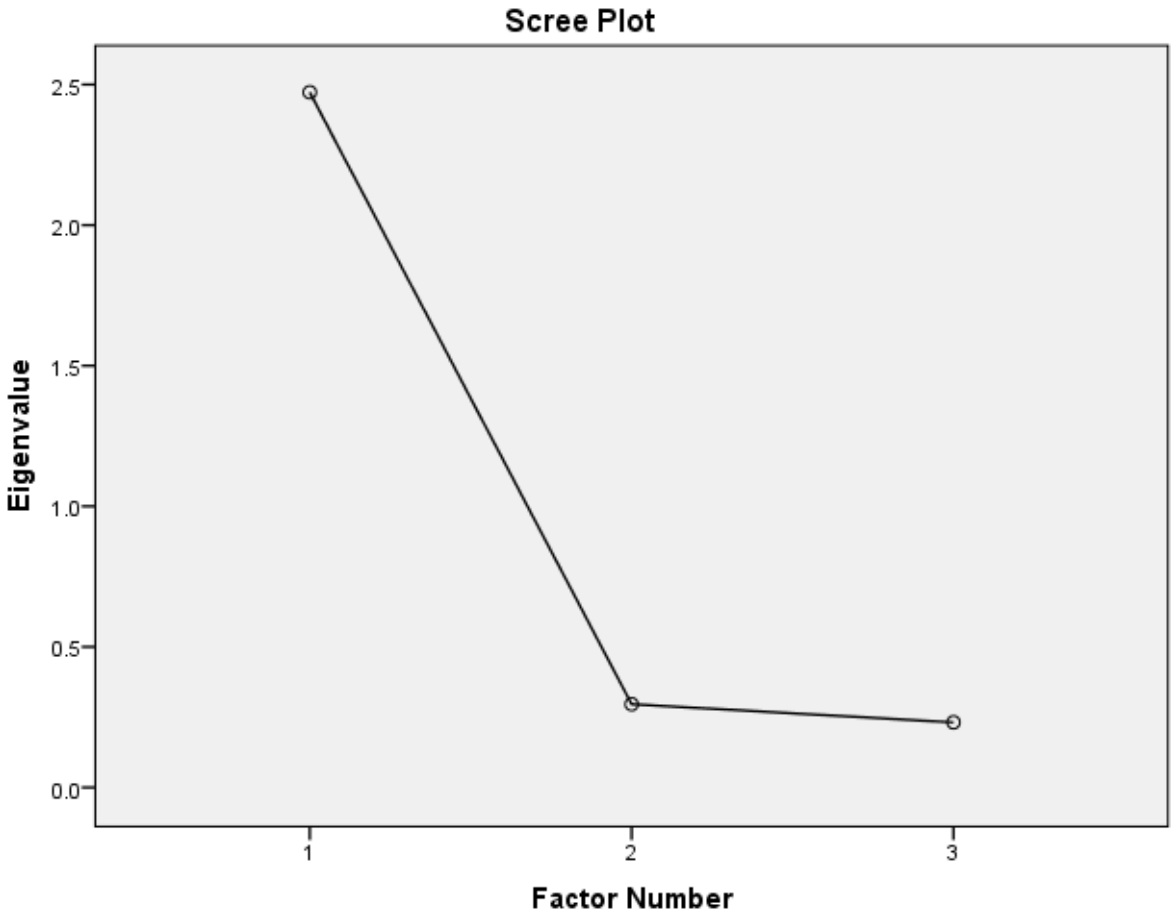


The analysis was then repeated to create a latent variable of "Low back pain". In this analysis, the software package selected 588 cases for analysis. This was verified as appropriate by the KMO statistic (0.746) and Bartlett's test of sphericity was also significant, indicating that the correlations between each scale were adequate ($X^2_{(3)}$ = 1038.707, p<0.05). The factor analysis showed that the three measures explain 74% of the variance for the generated factor (confirmed by the scree plot (Figure 6.7)), and the strong correlations in Table 6.12 below indicate that the three measures represent the extracted factor well.

Table 6.12 Factor matrix (EFA) for the low back pain latent variable: One-factor extraction model, entire STarT Back population (n=588)

Measure	**Factor loading**
Residualised change in Average Pain	0.90
Residualised change in Current pain	0.84
Residualised change in Least pain	0.84

Figure 6.7 Scree plot (EFA) for the low back pain intensity latent variable: One -factor extraction, entire STarT Back population



The two latent variables were then both included in a CFA model to test for co-variance between them. The model created in AMOS below (Figure 6.8) indicated a high correlation of 0.70 between the latent variables of low back pain and psychological distress. The model fit statistics (Table 6.13) do not clearly indicate whether this model is a good fit to the data. The CMIN/DF, CFI and SRMR values all indicate good or adequate model fit, while the RMSEA and X^2 (CMIN) do not. As noted above, the goodness-of-fit values and the judgement made based upon them is subjective. As the majority of the fit indices indicated good fit, the judgement here was that the model provides an adequate fit to the data.

Figure 6.8 CFA of co-variation between psychological distress and low back pain

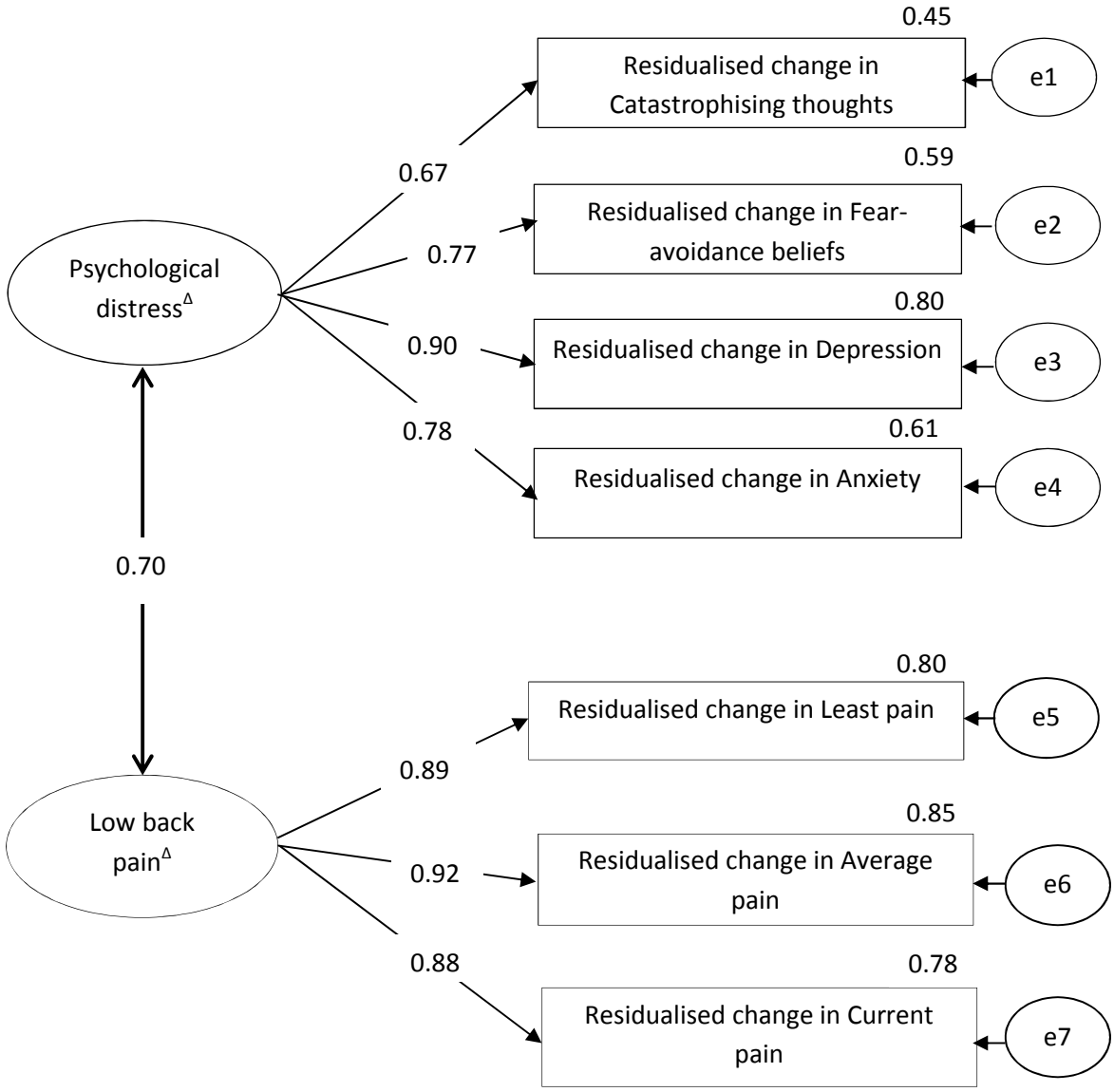


Table 6.13 Model fit statistics for the relationship between the latent variables of psychological distress and low back pain

Model Index	**Current Model**	**Good Model Fit**
CMIN*	41.81	Non-significant result
DF	13	
P	0.00	
CMIN/DF	3.22	Between 2-5
CFI	0.96	>0.95
RMSEA	0.13 (0.09 to 0.17), PCLOSE 0.00	<0.08
SRMR	0.06	<0.08

Because of the strong relationship between the latent variables, an EFA was performed to see whether the observed pain measures were also represented by the psychological distress factor. The EFA was carried out on the whole STarT Back dataset, using the ML method and direct oblimin rotation. 561 cases were included in the analysis. The adequacy of this was verified by the KMO statistic (0.85). Bartlett's test of sphericity indicated that the correlations between the scores were high enough for the analysis to be conducted ($X^2_{(21)} = 1953.78$, $p<0.01$). No set number of factors was requested for extraction from this dataset. The analysis extracted two factors, which together explained 61% of the variance. The rotated pattern matrix (Table 6.14) clearly indicates two separate factors, with the pain measures separate from the original measures of psychological distress. This analysis demonstrated that despite the high correlation between the latent variables, the observed measures of low back pain are still distinct from the measures used for psychological distress and were therefore treated as two separate factors in the mediation analysis.

Table 6.14 Factor matrix (EFA) including measures of pain as well as psychological distress: Direct oblimin rotation

Measure	Factor loadings	
	Factor 1	**Factor 2**
Residualised change in Average pain	0.92	
Residualised change in Least pain	0.84	
Residualised change in Current pain	0.80	
Residualised change in Depression (HADS-D)		0.84
Residualised change in Anxiety (HADS-A)		0.79
Residualised change in Fear-avoidance beliefs (TSK)	0.12	0.50
Residualised change in Catastrophising thoughts (PCS)	0.23	0.49

This pre-mediation analysis showed that each of the four psychological variables that were found to meet the criteria for being potential mediators all loaded onto a single psychological distress factor in an EFA analysis, which was confirmed by a CFA analysis on a randomly selected half of the STarT Back dataset. This supports the hypothesis underlying the STarT Back intervention that psychological variables may represent facets of an overall measure of psychological distress. The three measures of pain were found to represent a single latent factor of low back pain. Although the latent variables of psychological distress and low back pain were found to be highly correlated, when an EFA was performed on all of the measures together the pain measures loaded onto a separate factor from the psychological distress measures. This suggests that although the factors are related, they are measuring different constructs. In the mediation analysis therefore, an arrow to represent co-variation between the two latent variables was included in the SEM models.

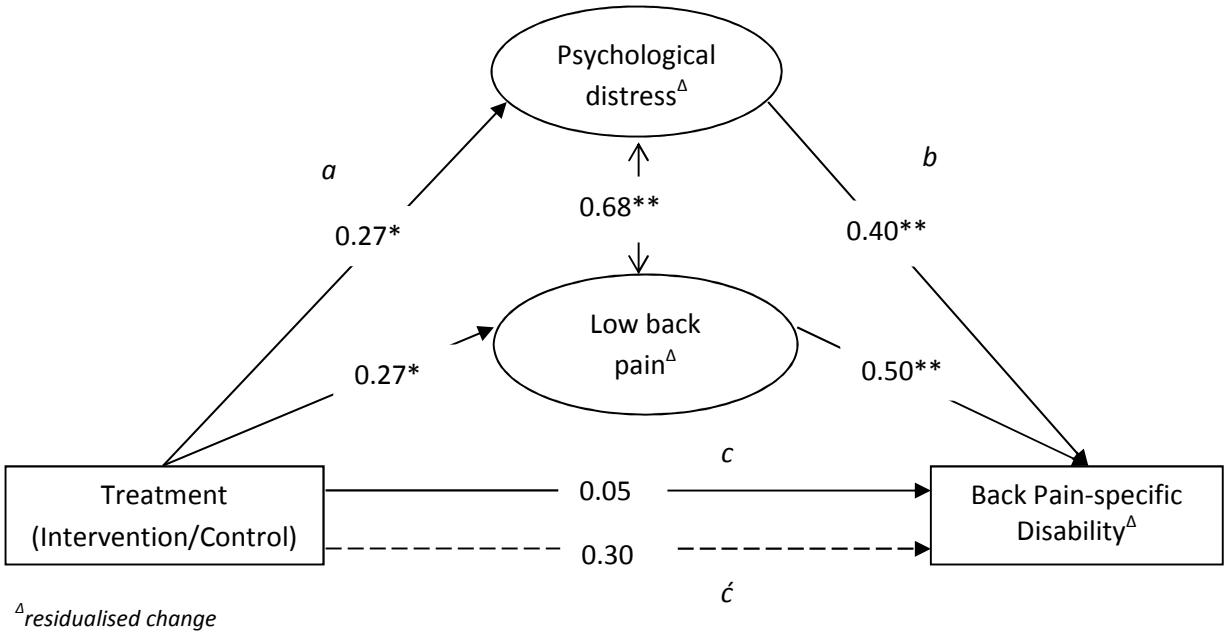
6.6.3 Mediation analysis

The mediation model with residualised change in disability (RMDQ) as an outcome is shown below (Figure 6.9) for the high-risk group. The statistics of model fit (Table 6.15b) suggests that this model, with co-variation between the two latent variables, provides adequate fit to the data although the RMSEA and X^2 indicate inadequate fit. When a binary variable is included in the model (in this case, treatment allocation), the value between this variable and the next variable in the pathway is denoted as an average treatment effect, or the effectiveness of treatment compared to outcome (Emsley et al 2010) (if control is allocated a value of 0 and treatment a value of 1, as is the case here). For example, in Figure 6.9, change in psychological distress at four-month follow-up is 0.27 standardised units higher for the intervention arm than for the control arm, suggesting that more change occurred in the mediator variable in the treatment group than in the control group. The effects of the model (see Table 6.15a) show that the total effect of the model is 0.30. As described in Chapter 1, this total effect can be broken down into the direct and indirect effects, with the direct effect being the effect of treatment on outcome and the indirect effect being the effect of treatment on outcome through the mediating pathway. The direct effect of treatment on outcome appears to show only a small, non-significant treatment effect of 0.05, because once the mediator variables (combined for psychological distress and low back pain) were added to the model the pathways of these variables explained a considerable proportion of the treatment effect; a statistically significant indirect (mediating) effect of change in the latent variables was found for the relationship between treatment allocation and change in disability (standardised indirect effect 0.25, bootstrapped 95% CI 0.09 to 0.39).

The values for each path in the model indicated that both change in psychological distress and change in low back pain have strong, significant relationships with change in disability, and a weaker but similar relationship was found between treatment allocation and change in the

psychological mediators. The standardised beta values for the association of the back pain latent variable with outcome are larger than for the psychological distress variable, suggesting that pain intensity explains more of the treatment effect than the psychological factors.

Figure 6.9 Full SEM model for mediating effect of changes in psychological distress and pain intensity on change in disability: High-risk group



$^{\Delta}$*residualised change*
**p<0.05*
***p<0.01*

All values are standardised

Table 6.15a Total, direct and indirect effects of the mediation model on change in disability for high-risk patients

	Effect	**Model**	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ$^{\Delta}$	Total (*ć*)	0.30 (0.14 to 0.43)	0.63 (0.29 to 0.94)
	Direct (*c*)	0.05 (-0.05 to 0.16)	0.11 (-0.11 to 0.33)
	Indirect (*ab*)	0.25 (0.09 to 0.39)	0.52 (0.19 to 0.85)

$^{\Delta}$*residualised change*

Table 6.15b Model fit statistics for mediation model of change in disability for high-risk patients

Model Index	Model	Good Model Fit
CMIN*	54.36	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.36	Between 2-5
CFI	0.96	>0.95
RMSEA	0.10 (0.07 to 0.13), PCLOSE .01	<0.08
SRMR	0.05	<0.08

Appendix 6.3 provides a comparison of the model presented in Figure 6.9 above (complete case analysis) and an analysis using all available data. This analysis used Full Information Maximum Likelihood (FIML), which is a missing data procedure that enables the use of all available data in the dataset to estimate values to be used in the analysis. This sensitivity analysis resulted in much weaker coefficients on the *a* and *c* paths than when complete case data only was used, and a very small total effect was found. Subsequent non-response bias to investigate this (Appendix 6.7) suggested that there were differences between responders and non-responders at four-month follow-up in terms of age (responders were older), baseline values of the outcome measure (responders were less disabled) and baseline scores on the psychological measures (responders were less fear-avoidant, catastrophised less and were less depressed). These differences suggest attrition bias in the complete case data sample, which is likely to affect the results. However, the FIML analysis assumes data are missing at random, and the estimates presented in Appendix 6.3 may still therefore contain bias.

Appendix 6.5 provides the mediation analysis using raw change scores, rather than the residualised change scores presented above. This is because raw change scores are more intuitive than residualised change, which give much smaller values for analysis. The analysis of raw change

showed that the coefficients are similar, and while the effects do differ in terms of the values provided the interpretation of them is the same.

When using general physical function (SF-12 PCS) as the outcome the indirect effects indicated a weaker but statistically significant mediating effect of change in psychological distress and low back pain on change in outcome (standardised indirect effect -0.21, 95% CI -0.34 to -0.08), and here as well, the association between change in pain with change in function appeared to be stronger than for change in psychological distress. The model fit statistics were similar to those for low back pain-specific disability change, suggesting that overall the model provided an adequate fit to the data (see Appendix 6.1).

The models for the medium- and low-risk groups supported the hypothesis that less variance would be explained by the latent variables in these groups, and the associations were generally weaker along each pathway compared to the high-risk models. As hypothesised the associations were weakest in the low-risk subgroup, as this group had the lowest scores on the psychological factors in the STarT Back tool and the targeted low-risk intervention was not designed to specifically address psychological distress (Appendix 6.2).

6.7 Discussion

6.7.1 Summary

The STarT Back trial was originally designed to test whether a model of stratified care consisting of targeted treatment matched to prognostic subgroups would lead to improved patient outcomes such as self-reported disability compared to best current primary care for back pain. The aim of the secondary analysis in the present study was to test the hypothesis that the observed

favourable outcomes of the stratified intervention in the STarT Back high-risk group were mediated by changes in psychological distress. It was hypothesised that the strongest mediating effects would be found in the high-risk group, as patients in this group were the most distressed and (if allocated to the intervention arm) received treatment designed to target this psychological distress.

This secondary analysis involved a series of preliminary analyses before the mediation analysis itself could be carried out. To summarise, the preliminary analyses described in Sections 6.5 and 6.6 above demonstrated that four of the psychological variables used in the STarT Back trial met the criteria as potential mediators of outcome (that is, they changed over time and this change was associated with change in outcome for each potential mediator variable, and treatment allocation was also associated with a change in each potential mediator). Alternative explanations for any mediating effects were also explored, and residualised change in pain intensity emerged as a possible alternative candidate construct to include in the mediation analysis. Once it was clear which variables were going to be included in the mediation analysis, the next step was to see if latent variables could be successfully created, which was demonstrated using EFA and CFA (measurement model) for the potential mediator (psychological distress) and the alternative mediation explanation (low back pain). Finally, the SEM models were run to test whether the relationship between patients' allocation to the high-risk treatment group in the STarT Back trial and change in their score on the disability and general physical function outcome measures (i.e. the effect of the STarT Back intervention on physical function outcomes) was mediated by either psychological distress or low back pain (structural model). For the high-risk analysis, psychological distress and low back pain did have statistically significant indirect (mediating) effects on both functional outcome measures and the hypothesis that this effect would be weaker in the medium- and low-risk groups was supported.

6.7.2 Comparison with conceptual model

Overall, the results of the analyses provide some support for the conceptual model proposed. The factor analysis performed on the four psychological measures suggested that they all correlated highly with each other and loaded onto a single factor of psychological distress. Also, the psychological variables did have a stronger mediating effect in the models where patients were classified as high-risk and went on to receive the high-risk treatment (if allocated to the treatment arm), compared to the analysis in the medium- and low-risk groups. Action and conceptual theory described in Chapter 1 would describe the stronger associations between psychological distress and disability (conceptual theory) as showing that targeting psychological distress was important, but that the weaker associations between treatment allocation and psychological distress (action theory) suggest that the psychologically informed physiotherapy did not target these factors as strongly as they could have done. This possibly reflects the fact that this treatment, although targeting psychological factors, did not target them as specifically as a more intensive treatment delivered by psychologists, such as CBT.

The alternative explanation to change in psychological distress tested in this study, change in low back pain intensity, was found to be more strongly associated with disability outcome and was a stronger mediator of the relationship between pain and disability than psychological distress. The two latent variables of psychological distress and low back pain were also found to have a strong relationship with each other. However, change in the variables was analysed as occurring in parallel rather than testing whether improvement in one leads to improvement in the other. The focus of this study was to test whether the trial authors' theory of how the trial worked was correct rather than test a more complex model of multiple mediating factors. The goodness-of-fit statistics provided for the mediation model presented above did not all meet the criteria for good fit, indicating that the hypothesised pathways may not be the only explanation for the effect of

treatment on change in disability. Future research could test a more complicated model which includes more potential mediators in a single pathway, to show a process of change in several variables as part of the treatment process. This would also allow testing of other variables that might be important in leading to a change in outcome, such as treatment expectations, that may help to further explain how the treatment worked.

Fritz et al (2012) suggest that in terms of power, it may be better to focus on increasing the strength of the conceptual theory, or focusing on the link between the mediator and the outcome, than it is to focus on trying to improve the action theory, or how the intervention can affect the mediator. They state that this may be most important in studies with small sample sizes. However, it has been suggested that the strength of the associations of the various paths has an effect on the study's power to detect a mediating effect, with a larger '*a*' path resulting in a larger sample size needed to test the '*b*' and '*c*' paths, due to more variance being explained by the intervention and therefore less variance available to be explained by the other paths (Frazier et al 2004). In this study, the sample size was not large because only the high-risk group was focused on and complete case analysis was used. It could therefore be that the smaller *a* path seen in this study is a result of the small sample size of the study. However, when comparisons were made between the analysis with all available data and the complete case analysis (Appendix 6.3) it appeared that the coefficients on the *a* path were larger in the complete case analysis, suggesting that sample size does not explain the smaller coefficients seen.

6.7.3 Comparison with previous findings

Few studies of treatment mediation in MSK pain populations have been conducted to date. The systematic review conducted as part of this thesis (see Chapter 4) located only 10 studies, and few of these had conducted mediation analysis according to the 'gold standard' as currently reported

in the literature. Despite the limitations of the analysis conducted here (see Section 6.7.4), it adds important information to the current evidence base. The indirect (mediating) effects for the high-risk STarT Back intervention appear weaker than those reported in studies included in the systematic review (Chapter 4) (standardised indirect effect of 0.25 reported in Table 6.17a; -3.82 reported by Robinson et al 2013; -0.23 to -1.18 reported by Turner et al 2007; -3.65 reported by Wicksell et al 2013), but these studies used more intensive therapies, such as CBT and ACT. The STarT Back intervention was deliberately designed to be a less intensive "light-touch" therapy that could still address psychological factors. The results found in this analysis therefore make sense, in that less intense psychological therapy will have a smaller indirect effect on outcome than studies with a more intensive intervention. It could also be that the way change was examined in each study had an impact on the strength of the findings. The present study used residualised change scores whereas Wicksell et al (2010) and Turner et al (2007) used either post-intervention scores or raw change scores. It is likely that stronger effects would be found with the methods used by other authors, as the calculated change would be greater. However, the purpose of mediation analysis is to specifically look at whether change in factors targeted by the intervention led to the change seen in the outcome. Only examining change that occurred as a result of treatment is therefore important. Residualised change scores are uncorrelated with baseline scores (Gillespie et al 1994) which are known to be strong predictors of future outcome, so the results presented here only show change that is not accounted for by baseline scores (Gillespie et al 1994). While in an RCT differences between intervention groups in baseline scores are accounted for, from a mediation perspective this only holds for paths *a* and *c*, and not for the *b* path as the mediator score is an outcome of randomisation (see Chapter 3 for a detailed discussion of this). The use of residualised change has been criticised for being difficult to interpret (Wurm & Fisicaro 2014), but a sensitivity analysis performed on this data using raw change scores (high-risk group only, complete case analysis) found that the results and their interpretation were similar to that using residualised change (see Appendix 6.5).

6.7.4 Potential limitations

The FIML (sensitivity) analysis and subsequent non-response bias analysis (presented in Appendices 6.3 and 6.7) suggested that the participants included in the complete case analysis differed substantially from those who were excluded. This was particularly apparent in the psychological measures, where responders at four-month follow-up had lower baseline scores of fear-avoidance beliefs, catastrophising and depression compared to non-responders. It is known that participants who had higher fear-avoidance beliefs at baseline were more likely to drop out of the control group if their fear-avoidance beliefs were not being managed effectively (J Hill, personal communication) but this suggests that people with higher scores on several key factors in both arms dropped out if their baseline scores were higher. This attrition bias means that the results presented above need to be interpreted with caution, as they represent only a subsample of the high-risk population included in the trial.

The psychometric properties of measures to be used for mediation analysis are important to assess to ensure the measures are suitable for inclusion in the analysis (see Chapter 3 Section 3.3.3). In this study, only one property (internal consistency) was formally tested. The importance of this in relation to mediation analysis is that we need to be confident that the tool we are using to measure a construct is an accurate representation of the construct we believe to be mediating outcome, and that this measure can detect change over time. For the TSK and PCS, reviews of psychometric properties exist and show support for these tools being responsive (Woby et al 2005) and having good test-retest reliability (Sullivan et al 1995). The PCS has three subscales of Rumination, Magnification and Helplessness (Sullivan et al 1995) but can also be used as a total score (Severeijns et al (2002) reported a similar fit for the three- and one-factor structures in LBP patients). As the scale was used as a single measure in the STarT Back trial it was used as such in the present study also.

The HADS has less evidence for its reliability and validity in MSK pain populations, with mixed evidence for its ability to predict disability (see Chapter 5). However, as the HADS was the only measure of anxiety and depression available in this dataset, and items from this tool were used in the STarT Back tool, this measure was included in the present analysis. A related point is that the Cronbach's alpha value in the internal consistency analysis for the Pain Catastrophising Scale was very high, suggesting potential item redundancy. However, the purpose of the analysis was to include and investigate all of the factors measured in the STarT Back tool, as these were the factors which categorised patients as high-risk. Future analysis could examine whether the catastrophising measure is needed.

Responsiveness may also be important to assess in mediation studies to ensure the measures can adequately show minimal important change, although sensitivity to change was assessed in this study as part of the analysis by investigating associations between residualised change in the mediator and residualised change in outcome, suggesting that this is adequate for each of the measures included in the analysis. However, the mediation analysis focused on latent variables rather than these individual constructs. One of the strengths of SEM is its ability to include several measures of a construct by creating latent variables. The factor analyses conducted as part of the preliminary analysis (EFA and CFA) showed that the variables used to create the latent constructs of psychological distress and pain intensity loaded strongly onto the constructs and provided stronger measures than each of the measures individually. These analyses therefore demonstrated convergent and discriminant validity of these two constructs, which assess whether the items of interest represent the same construct and whether the constructs can be distinguished from each other. Using additional measures may ensure the construct is measured with more accuracy and therefore increase construct validity.

All participants in the STarT Back trial, including those in the control arm, received half an hour with a physiotherapist prior to them commencing treatment. This could have had a significant impact on the outcome of this study. The time could have been very beneficial for the patients in the control group and answered any queries or worries they may have had, which could have reduced their scores on the distress measures and improved the effect of the control intervention. This means that the difference between the two treatment arms may have been smaller than what would have been expected had a control arm receiving no treatment or usual primary care been used.

Also, patients were not categorised based on their pain duration. Different levels of pain duration may have affected patients in different ways – patients with new pain episodes are likely to have lower levels of catastrophising thoughts or fear-avoidance beliefs, while patients with persistent pain may have higher levels of distress. One way of testing this would be to carry out a moderated mediation analysis (e.g. Baron & Kenny 1986). However, it was felt that subgrouping patients further when some of the risk subgroups only contained small numbers would have affected the reliability and robustness of the analysis. Pain duration was examined as a possible confounding factor in Section 6.6.1, and found to have only weak, if statistically significant, relationships with the outcome measures used in this study.

It is possible that the number of follow-up points and the timings of when these follow-ups took place may have had an important influence on estimation of the mediating effects. In the STarT Back trial, as with many intervention studies, patients completed measures at baseline and then post-intervention. It is likely that mediating effects were missed because not enough measurements were taken from participants, or perhaps because no measurements were taken while the intervention was being delivered. It is important to acknowledge that the STarT Back

trial was not designed to test mediating effects. The issues discussed above could be resolved by designing trials to directly assess mediation from the outset with optimal time points for assessment, rather than conducting *post hoc* analyses.

Related to this is the fact that temporality cannot be established in this study, even though the data is taken from an RCT. This is because the analysis presented, while complex, still only shows change between the mediator and outcome variables at the same time points. It therefore cannot be established whether the hypothesised order of the variables as presented in the models above is correct. Additional measurements of all variables of interest would therefore help to establish when change occurs and the order in which it occurs.

It is also unclear whether the right constructs were chosen for the analysis, as many other factors, both psychological and non-psychological, could also be on the mediating pathway. For example, it is likely that other factors such as clinician confidence, self-efficacy, or social factors such as emotional support play an important role but are often not measured or are difficult to measure within a trial. As mentioned in section 6.7.2, the goodness-of-fit statistics generated for the mediation model presented in this chapter suggest that fit was not optimal and that it is likely that there are other unmeasured mediating pathways that could be explaining the treatment effect. However, the psychological measures included in this analysis were chosen because of the strength of the evidence suggesting their prognostic value, which was gained from the available literature and also from the analysis presented in Chapter 5. The preliminary analysis presented in both Chapter 5 and the present chapter also showed how the variables included in this analysis met criteria for mediation prior to being included in the mediation analysis.

Another point for consideration is whether a different method of mediation analysis may have resulted in a different conclusion. Chapter 3 gives an overview of the different approaches to mediation analysis available. SEM was chosen because it has several strengths over other methods, but it is a complex form of analysis and many preliminary steps must be taken before it can be conducted. An alternative method of mediation analysis, such as linear regression, may have given different results. However, the point of this analysis was not to compare different methods of mediation analysis, as this has been done previously (MacKinnon et al 2002; Hayes & Scharkow 2013). Instead, the aim was to use the most appropriate method to answer the hypothesis under investigation. In STarT Back, the hypothesis was that psychological distress (rather than a single psychological construct such as depression) would be reduced by the psychologically informed intervention offered to the high-risk group, leading to the reduction in disability. Therefore SEM, with its ability to assess latent variables, was the most appropriate method to use in this study. Alternative methods such as linear regression or latent growth modelling were not possible to conduct as a comparison because they would not have been able to account for the latent factor of psychological distress.

The analysis presented in this chapter set out to capture a broad concept of psychological distress rather than the more specific constructs (e.g. fear-avoidance beliefs, anxiety) that were used to measure this latent variable. Such composite variables have been found to show mediating effects in previous studies (e.g. Wegener et al 2011) and while each *construct* is distinct, it is likely that self-report measures of psychological factors are related. Indeed, the preliminary analysis presented in this chapter found that changes in the psychological variables tested were very strongly correlated with each other. The use of a latent variable rather than individual psychological measures also reflects the conceptual model of the STarT Back trial, which did not aim to assess individual constructs. This sets the trial apart from other psychological treatments for LBP in primary care, such as the Progressive Goal Attainment Program (PGAP) approach (e.g.

Sullivan & Simon 2012) and the Back In Action approach (Von Korff et al 2005) which also target psychosocial barriers to recovery, but focus more specifically on fear-avoidance beliefs. The Von Korff study is analysed in more detail in Chapter 7.

6.8 Conclusion

The psychological variables that physiotherapists attempted to manipulate during the STarT Back high-risk intervention explained part of the outcome effectiveness seen in the trial. The BeBACK analysis presented in Chapter 5 provided a strong empirical basis for the selection of the variables included in the analysis presented here, highlighting the value of observational study data in helping choose appropriate potential mediators of treatment effect. The present analysis also emphasises the importance for intervention studies to be underpinned by a clear theoretical or conceptual model. However, this analysis only looked at change between two time points which was not enough to assess the temporal order of the variables in the model. Trials that are designed to adequately test for mediating effects, with variables being measured at appropriate time points which fit in with that model, are required.

The next chapter builds on the analysis outlined here, and tries to address some of the issues and limitations that could have affected the mediation analysis of the STarT Back trial. It will also test an alternative method of analysis, recommended in the literature summarised in Chapter 3, which extends SEM to include more assessment time points.

Chapter 7: Using RCT Data to Investigate Potential Mediators of Treatment Outcome

Part II: Back In Action trial

7.1 Introduction

This chapter builds on the previous chapter which presented a mediation analysis of a randomised controlled trial (RCT) investigating a stratified care approach for low back pain (LBP) including a psychologically-informed intervention for those at a high risk of poor outcome (STarT Back). The present chapter will also outline analysis of an RCT, which is arguably more suited to a mediation analysis because the intervention had a more explicit psychological conceptual framework underpinning the mechanism by which the active treatment was thought to lead to improvements in functional outcomes. In addition, in this RCT there were more assessments made over time, and the intervention itself was more specifically focused on a key potential psychological mediator (fear-avoidance beliefs). This specificity of focus on a core mechanism facilitates consideration of mediation. This chapter investigates longitudinal associations between the potential mediator and outcome variables through the use of latent growth curve modelling (LGM) in order to carry out mediation analysis.

7.2 Rationale for Study

Temporality (change over time) has previously been identified in this thesis as being fundamental in establishing whether a proposed mediator of treatment outcome is on a causal pathway. The review in Chapter 3 highlighted that in order to truly establish how changes in factors are related over time, at least three measurements should be taken over different time points (e.g. Kazdin

2007). Typically in intervention studies, measures are only taken pre- and post-intervention which only gives information on whether a change takes place or not, and gives little information about how a factor changes over time as a result of the intervention. The purpose of this chapter was therefore to investigate a way of analysing intervention study data where data has been collected at multiple time points.

The Back In Action study (Von Korff et al 2005) compared a psychological intervention with treatment as usual in 240 patients with LBP. The intervention centred on addressing patient's fear-avoidance beliefs and helping to improve their functioning, and it was hypothesised by the trial authors that this would lead to a reduction in fear-avoidance beliefs and improved functional outcomes. Measures of psychological factors and function were taken at baseline and four follow-up points, and incorporating all of these into the mediation analysis provided more information on longitudinal associations between psychological factors and disability than just looking at changes in mediators and outcomes between two of those time points. The present mediation analysis investigated fear-avoidance beliefs as a potential mediator variable of self-reported disability.

It should be stated at this point that although five measurement points were available for analysis in the present study, none of these were made during the intervention period. This means that the issue of temporality and therefore causality, as discussed in detail in Chapter 3, still cannot be truly established for changes occurring during the treatment period. However, the purpose of this thesis as a whole is to investigate how best to conduct mediation analysis and provide guidance as to how to design future studies to best test for mediators of treatment effect. Therefore, while the analysis presented in this chapter cannot confirm a mediating pathway informed by repeated assessments during the course of treatment, it can provide evidence for a particular analysis

technique which could potentially be useful in future studies. Latent growth modelling (LGM) is a technique which allows the inclusion of multiple time points, thereby providing a more accurate estimate of change (Byrne et al 2008). It has been recommended that analyses using LGM requires a sample size of at least 200 (Byrne et al 2008), and the Back In Action study meets this criterion.

7.3 Back In Action Trial

The Back In Action trial was an RCT designed to compare an intervention addressing fear-avoidance beliefs with treatment as usual for chronic LBP in order to improve patient function (as estimated by reduction in self-reported disability). The trial recruited 240 adult participants (119 in the treatment group and 121 in the control group) aged between 25 and 64 years suffering with chronic LBP from primary care settings in Seattle, Washington State, US. Assessments took place at baseline and then post-intervention at two-, six-, 12- and 24-month follow-up. The numbers of males and females and the levels of education, pain intensity, fear-avoidance beliefs and disability were similar across both groups at baseline, indicating successful randomisation (see Table 7.1).

Table 7.1 Baseline characteristics of study population (*n*=240)

Variable		**Treatment (*n*=119)**	**Control (*n*=121)**
Age in years, mean (SD)		49.67 (9.02)	49.82 (9.77)
Sex, % female		64.7%	60.3%
Education in years*, median (range)		6.00 (3-8)	6.00 (3-8)
Pain Intensity (1-10), mean (SD)	Average last three months	5.71 (1.84)	5.83 (1.82)
	Today	3.45 (2.49)	3.45 (2.57)
Pain duration (days), mean (SD)		110.50 (60.72)	97.39 (62.34)
Fear-avoidance beliefs (TSK – 10 item version), mean (SD)		41.47 (8.79)	41.25 (8.22)
Disability (RMDQ-23 item version), mean (SD)		12.10 (5.49)	11.36 (5.67)

**Education was graded in eight levels by asking participants what the highest grade of year of school they completed. This ranged from 1 (less than eight years) to 8 (postgraduate)*

The overall aim of the original study was to test whether the proposed model of 'activation' (addressing fears and encouraging resumption of normal activities (Von Korff et al 2005)) could feasibly be employed in primary care. A secondary aim was to assess whether the proposed intervention led to a reduction in fear-avoidance beliefs and disability compared to usual care. It was this secondary aim that was of interest in the present study, as this related to how the authors proposed the trial was working.

7.3.1 Intervention: Conceptual model

The intervention used in the study was based on a suggested framework by Balderson et al (2004) of assessing and addressing a patient's fear-avoidance beliefs, anxiety and activity limitations, performing a clinical assessment to rule out red flags (physical signs suggestive of the need for an urgent surgical opinion) and providing treatment recommendations in line with goals set by the patient. The intervention involved four face-to-face visits, two with a psychologist and two with a physiotherapist, with the first session focusing on goal-setting and putting together an action plan

and the remaining sessions focusing on how to achieve those goals and addressing barriers to achieving the action plan. The three physiotherapists involved in the study received half a day formal training from the study psychologist but most of the training was informal (B Balderson, personal communication) and no formal evaluation of the training was recorded. A summary of the intervention components is given in Box 7.1. In addition to these visits, patients also received a book on back pain self-management (Moore et al 1999, cited in Von Korff et al 2005) and a 40 minute video on back pain (Patient Education Media Inc. 1996, cited in Von Korff et al 2005). The control group received treatment as usual, which Von Korff et al acknowledged could be very diverse and included medication and physiotherapy sessions.

Box 7.1. Back In Action (BIA) intervention

Visit 1: Psychologist (90 minutes)

- Identified and addressed patient fears about pain
- Discussed resuming normal activities
- Set activity and exercise goals
- Developed action plan (in collaboration with the patient) on how to achieve those goals

Visit 2: Physiotherapist (60 minutes)

- Standardised physical examination of the back
- Discussion of any unresolved concerns from the first visit
- Taught stretches and exercises relevant to patient's goals and action plan
- Offered guidance in overcoming barriers to carrying out action plan

Visit 3: Physiotherapist (30 minutes)

- Focused on action plan and exercises relevant to that plan

Visit 4: Psychologist (20 minutes)

- Review of patient's progress
- Encouraged use of relaxation techniques
- Discussion of managing flare-ups and resuming normal activities when these occurred

7.3.2 Measures

The trial included a number of different outcome measures, with the primary outcome measure being low back pain-specific disability, measured using the Roland-Morris Disability Questionnaire (RMDQ). This was a modified, 23-item version of the original 24-item RMDQ in which five of the original items had been deleted and replaced by four other items reported to better detect changes in pain (Patrick et al 1995). The trial also measured fear-avoidance beliefs using a modified version of the Tampa Scale for Kinesiophobia (TSK) which contained 10 of the original 17 items. Von Korff et al (2005) stated in their paper that this tool was scored by multiplying the mean item score by 17 (the number of items in the original tool) so that the scores from the shortened version and the original version would be comparable. Both modified tools were still scored in the same direction as the original versions, with higher scores indicating higher disability or fear-avoidance beliefs. As described in Chapter 5, a score of 37 or above on the TSK has been documented as a cut-off for high fear (Vlaeyen et al 1995a; Bränström & Fahlström 2008) and a reduction in score by four points can be taken as representing a clinically significant improvement (Woby et al 2005). For the RMDQ a reduction in score of 30% has been found to show clinically significant improvement (Jordan et al 2006). Other outcome measures included back pain worry, pain intensity and mental and physical functioning.

7.3.3 Conclusions of original study

The intervention was successful in reducing disability across all time points compared with the control group, but the authors do acknowledge that the intervention group received more therapeutic contact than the control group, which could have enhanced the treatment effect along with other non-specific effects. They also state that the results cannot identify which part(s) of the intervention were most effective.

7.4 Aim

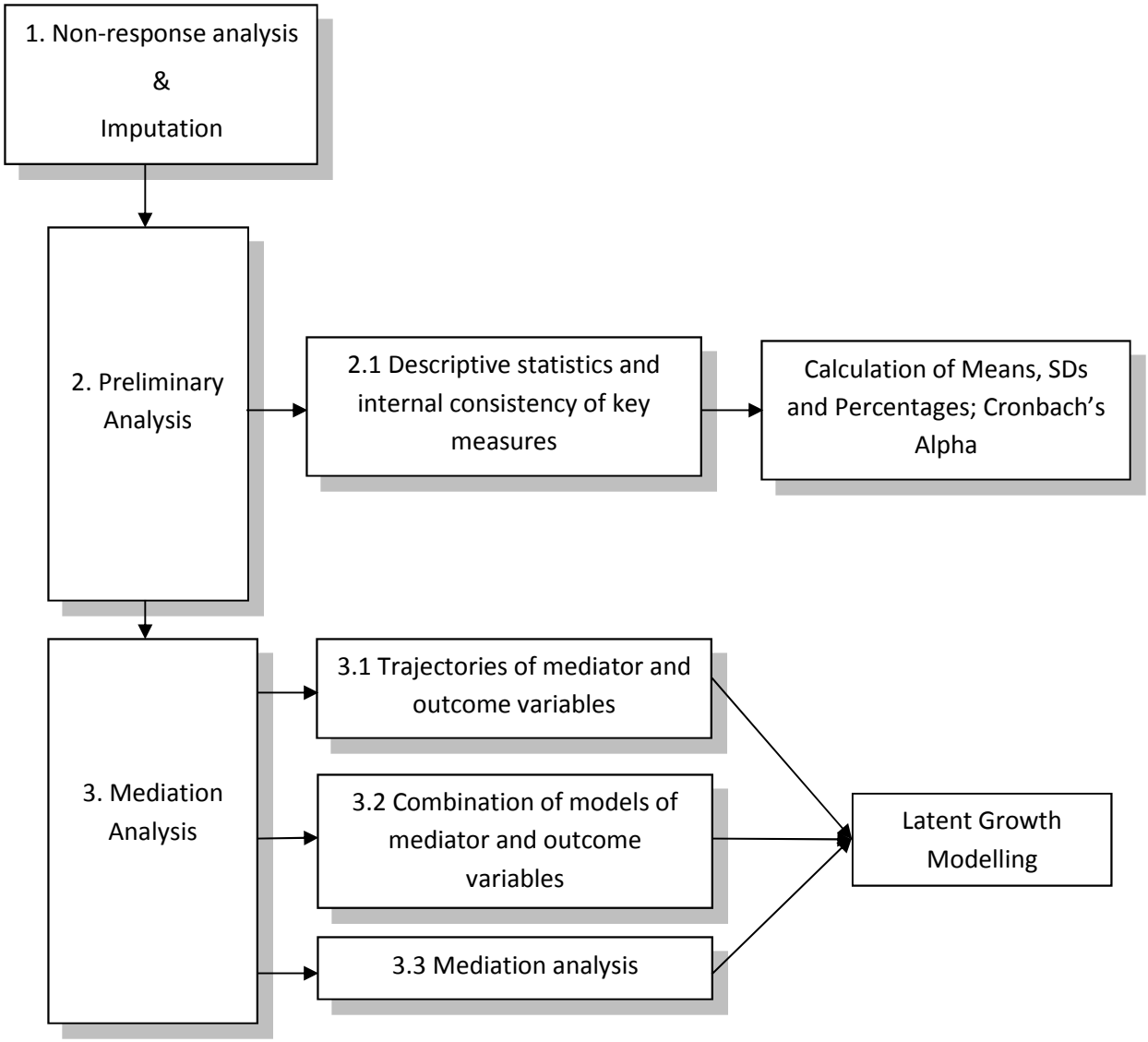
To test whether change in fear-avoidance beliefs is a mediator of the relationship between allocation to the active treatment intervention and change in disability.

7.5 Specific Objectives

1. To describe the baseline characteristics of the study population and the measurement properties of the mediator and outcome measures;
2. To examine the change in both the mediator (fear-avoidance beliefs) and outcome (disability) variables in the treatment and control groups and derive well-fitting LGMs for each variable;
3. To examine the effect of treatment group allocation on change in both the mediator and outcome variables;
4. To estimate the mediating effect of change in fear-avoidance beliefs on the relationship between the active intervention arm and change in disability.

All analyses were carried out using Microsoft Excel, SPSS PASW Version 18 and AMOS Version 19. A summary of the analyses performed is given in Figure 7.1.

Figure 7.1 Flowchart describing the steps of analysis



7.6 Methodology I: Imputation

7.6.1 Missing data

Missing data is a common occurrence in longitudinal data due to dropout, death or unavailability at follow-up (Collins 2006). If data is missing completely at random (MCAR), then the values of the missing data are not related to the values of the collected data (Enders et al 2001a) and are missing completely by chance (Sinharay et al 2001). This is unlikely to result in biased estimates and the remaining data can still be seen as representative. Data that is missing at random (MAR) is likely to be related to other observed variables (Sinharay et al 2001; Enders 2001b; Enders 2011). Data that are not missing at random (NMAR) have missing values that are related to the value that would have been seen had the information not been missing (Sinharay 2001), and are therefore most likely to result in biased estimates.

Several methods have been proposed to deal with missing data. Complete case analysis, where cases with missing values are simply removed from the analysis (as used in Chapter 5), is commonly used and simple to implement. However, current recommendations advise against this due to loss of power and potentially biased results (Collins 2006). Single imputation methods such as Last Observation Carried Forward or Mean Value Substitution help reduce the power problem, but do not account for any uncertainty or change in the data (Karahalios et al 2012), meaning the data is still likely to be biased. These approaches are therefore also not recommended (Li et al 2014). Some single imputation methods such as Maximum Likelihood (ML) estimation have been found to yield unbiased estimates if data is missing MCAR or MAR, although sensitivity analyses are recommended to check the stability of the results (Enders 2011). Multiple imputation (MI) (Rubin 1987) has recently been recommended as the most appropriate method of imputation (Olinsky et al 2003). This method involves replacing the missing values with a given number of alternative values, which are analysed separately and the estimates and standard errors are then

combined using Rubin's rules to give overall estimates for the results (Schafer & Olsen 1998; Sinharay et al 2001). This ensures that the results acknowledge the fact that there is some uncertainty about the true values of the missing data (Sinharay et al 2001). However, MI is complex and beyond the scope of this thesis. Expectation Maximisation (EM), a form of ML estimation, was therefore the method chosen to perform imputation in this analysis (see Section 7.6.2).

Whether or not the data is MCAR can be established via Little's MCAR test (Little 1988), which will be non-significant if the data is MCAR. However if the test is significant there is no way of establishing whether the data is then MAR or NMAR. This means a decision needs to be made as to whether the data from that time point are to be excluded as they may introduce bias, or included with the acknowledgement that some bias may be present. In addition to this test, baseline characteristics of responders and non-responders at each assessment point (split to show treatment and control groups separately) were also provided to examine differences between responders and non-responders at each time point. If, for example, patients who had higher scores on the TSK at baseline were more likely to drop out of the control group if their fear-avoidance beliefs did not greatly improve (which happened in the STarT Back trial, (J Hill, personal communication)), this suggests that there is a high risk of attrition bias.

7.6.2 Imputation

Imputation was necessary because of attrition over time in the follow-up assessments (Step 1 of Figure 7.1). By the 24-month follow-up point 79% of those who participated at baseline provided data, although participants were able to 'skip' a follow-up point and still be involved in the other assessments. This means that there is potentially a risk of bias if only cases with complete data are used for analysis. Incomplete data is restrictive when using SEM and LGM as certain goodness-

of-fit indices (i.e. Standardised Root Mean-square Residual (SRMR)) can only be used with complete data. Confidence intervals for the mediated effect, a requirement for studies of mediation (e.g. MacKinnon et al 2004; Hayes 2009) can also only be obtained on data with no missing values. As explained in Section 7.7.1 above, while complete case analysis allows us to analyse 'real' data collected during the study, it lacks power and potentially suffers from bias because of selective missing data (Vergouw et al 2012). Imputation was therefore carried out, but in line with current recommendations (e.g. Powney et al 2014) the complete case analysis is included in Appendix 7.1 for comparison.

EM was used to impute the data. EM involves predicting the missing values from the assumed parameter values of the data model, and then using those predictions to improve the parameter estimates (Schafer & Olsen 1998). EM also takes people's characteristics into account when computing the estimates, resulting in less biased, more reliable estimates than the complete case data (Collins 2006). This process of prediction is repeated several times until the best estimates for the missing data are found for inclusion (Schafer & Olsen 1998). This differs from multiple imputation, where several different datasets are created and then merged; EM instead only provides a single imputed value from the process described above and does not include error in the generated estimate, so therefore should only be used with small amounts of missing data (Tabachnick & Fidell 2007). EM assumes that the data is normally distributed and the missing values are MAR (Newman 2003). Cheung (2007) reported potential problems using EM with LGM in that the standard errors were often underestimated, but in the present analysis this particular statistic is not of key interest. In this dataset, only small amounts of missing data were present (less than 30 people at each follow-up point in either treatment arm out of the 240 included at baseline), so it is likely that bias as a result of EM imputation will be minimal.

7.7 Methodology II: Preliminary Analysis

7.7.1 Descriptive statistics

Before conducting the mediation analysis, descriptive characteristics of each of the key variables were obtained for the treatment and control groups (Step 2.1 of Figure 7.1). As explained in Chapter 4 (Sections 4.5.2 and 4.6.1), the psychometric properties of the tools used in mediation analyses are important to consider – the measure should adequately measure what it is supposed to measure (validity) and the measure should be consistent and reliable over time. This is so that if a change (or no change) is found in a mediating relationship, we can be more certain that this is a true reflection of the data and not a consequence of problems with the tool used to measure the constructs of interest. However, testing each of these properties within each dataset is often not possible, and data from other studies of the same population is often relied upon to provide information on the validity and reliability of a tool within our population of interest. One property which can be tested is the internal consistency of the measures used in the analysis. It is important that the items in the scale are correlated, as if this is not the case, it might affect the interpretation of the results. Other properties of the measures of interest in this chapter not already reviewed in Chapters 5 or 6 were reviewed in the literature and are commented on in Section 7.10.2.

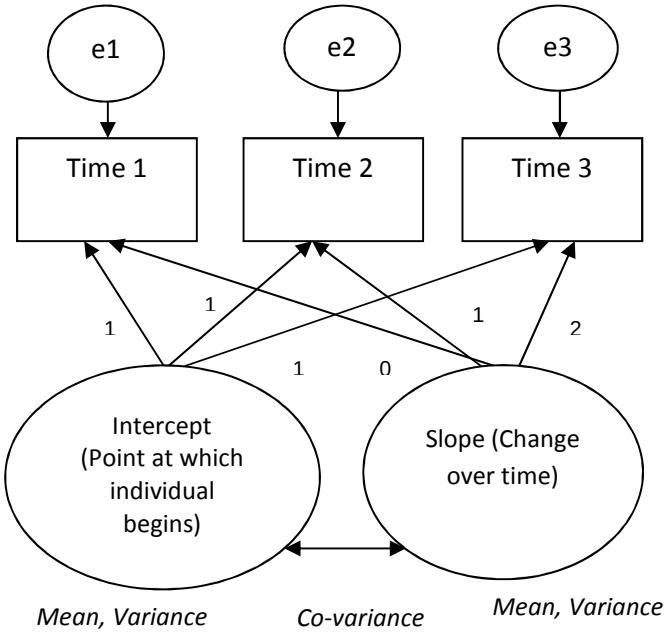
7.8 Methodology III: Mediation Analysis: Latent Growth Modelling

LGM allows the examination of variables across time, both within and between participants (Cheong et al 2003; Kosma et al 2012; Witkiewitz et al 2012). It can be carried out using SEM, like the previous mediation analyses in Chapter 6. This means advantage can be taken of the strengths of SEM, including the use of goodness-of-fit indices (explained fully in Chapter 6 and summarised below in Section 7.9.1).

LGMs consist of an intercept latent factor and a slope latent factor (Curran & Muthén 1999). The intercept factor represents the starting point for the trajectory of a factor (its baseline status) while the slope represents the growth (or change) in the trajectory of that factor over time. The use of the term "trajectory" in this context refers to the path each individual's fear-avoidance beliefs and disability scores take over the measured time points. Although in this analysis it is the slope that is the factor of interest, the intercept is also important as where a person starts in an intervention will have an impact on their scores at subsequent points (e.g. people who start with very high scores will have more room for improvement than people who had low baseline scores). Figure 7.2 shows a basic LGM with three time points to help illustrate how an LGM is put together.

The mediation analysis was carried out according to the steps set out by Cheong et al (2003) and Roesch et al (2009), described below, and also according to guidance by Byrne (2010) on testing for differences within models. Separate LGMs for the mediator and outcome were created to examine changes over time in each variable (Step 3.1 in Figure 7.1), and then both models were incorporated into a single model which examined whether the intervention was associated with a change in the observed trajectory of the mediator variable, and whether the observed trajectory of the potential mediator was associated with a change in the observed trajectory of the outcome variable (Step 3.2). LGM differs from Latent Class (Growth) Analysis, a technique which can also be used to model trajectories, because LGM uses the mean slope across all cases at each time point (Roesch et al 2009) rather than grouping people into different trajectories based on similarities in scores over multiple time-points.

Figure 7.2. Generic LGM with three time points



As in a standard SEM model, latent and measured variables and error are included in the model. However, in an LGM, the latent variables are the intercept and slope factors and the measured variables are each time point a variable was measured at. The numbers against each of the lines represent factor loadings, which are important in specifying the model. As the intercept is the point at which each individual starts, all the factor loadings from the intercept to each time point are set to the same value. Factor loadings for the slope factor need to represent both the trajectory of change in the measured variable over time and also reflect the time intervals between each of the measures (Cheong et al 2003). In Figure 7.2 above for example, the factor loadings of 0, 1 and 2 reflect linear growth at equally spaced intervals in time. The values chosen are arbitrary and not reflective of the actual change in scores on the measures included in the model.

LGMs can incorporate models where growth is not linear and where assessment intervals are not equally spaced, providing that all participants have been assessed at the same time points (Cheong et al 2003; Byrne 2010). It is also possible to allow the model to estimate certain factor

loadings when the trajectory of a measure is not clear (Cheong et al 2003), meaning that not all of the slope loadings need to be pre-specified before the model is run.

Both the intercept and slope factors also have a mean and variance value associated with them that are estimated from the data. In this case, the means represent the parameter values for the group, while the variances represent the individual variation of each person around the group means, so that larger variance suggests more individual difference in growth (Curran & Muthén 1999). In a standard LGM, co-variance between the intercept and slope factors is also included. This gives an indication of whether people who start at lower or higher scores on a measure also change at lower or higher rates.

7.8.1 Goodness-of-fit indices

Model fit indices show how well the hypothesised model fits the observed data (Schermelleh-Engel et al 2003). The goodness-of-fit indices used in LGMs are the same as those available in SEM, so the indices chosen for use in the previous chapter (Chi-square (CMIN), comparative fit index (CFI), root mean square error of approximation (RMSEA) and SRMR) were also used here. These indices are described in detail in Chapter 6 (section 6.5.4). However, although the CMIN is recommended to be reported as an indicator of model fit, it is often statistically significant in large samples (Byrne 2010) and may therefore not be an appropriate measure upon which to base judgements of goodness-of-fit (Schermelleh-Engel et al 2003; Byrne et al 2008; Williams et al 2009). Therefore, while it is reported here, the decision of whether or not a model provides a good fit to the data will be based on the other reported indices.

If the goodness-of-fit indices do not indicate good fit, steps can be taken to try and improve the model. It is recommended that the standardised residual co-variances of the model are inspected

for any values that exceed ± 1.96 (Schermelleh-Engel et al 2003), as this would indicate that the model shows variance values which are substantially different to the observed data, and so is not representative of the study sample. If any paths do exceed this value they could be allowed to freely estimate (i.e. the path is not given a factor loading) to try and improve model fit (Schermelleh-Engel et al 2003). Alternatively, modification indices can be examined to check whether the hypothesised model could be improved by adding or removing factor loadings on specified paths (Byrne 2010). Modification indices show the overall decrease in X^2 value that would occur if the factor loading for a particular path was not specified (Byrne 2010), with a large value indicating a larger improvement in model fit.

It is emphasised that changes to the original model should only be made if the modification makes theoretical sense and would substantially improve the model (Byrne 2010). For models that did not provide a good fit to the data in the present analysis, standardised residuals and modification indices were examined and both the original and modified model fit indices are provided for comparison.

7.8.2 Step 3.1: Investigate the shape of the observed growth trajectory for the mediator and outcome variables

This first step involved estimating growth trajectories, or change over time, for both the mediator and outcome variables separately from the observed data. The aim of this step was to ensure that the factor scores for the slope factor represented the observed trajectory (seen when plotting the mean values of each variable across the measurement points) and provided a good fit to the data. It also allowed examination of whether these trajectories were different in the treatment and control groups, which would be the case if the intervention had been effective (Cheong et al 2003).

As the Back In Action intervention had produced larger changes in the mediator and outcome variables in people in the intervention group compared to the control group, it was anticipated that the observed trajectories in each of these groups would differ. However, because the follow-up points were all measured after the intervention had finished (at two, six, 12 and 24 months) it was not expected that the decrease in fear-avoidance beliefs and disability scores would be linear over the entire follow-up period; it was considered more likely that an initial steep decrease between baseline and two-month follow-up would level out over time. This is shown in the original paper and in Table 7.2, which presents means for both variables at all measurement points in the treatment and control groups.

Table 7.2. Changes over time (means and SDs) for potential mediator and outcome variables in the Back In Action trial

Variable	**Treatment (*n*=119)**	**Baseline**	**Two-month follow-up**	**Six-month follow-up**	**12-month follow-up**	**24-month follow-up**
RMDQ (23-item version)		12.10 (5.49)	10.02 (6.27)	8.91 (6.60)	8.46 (6.90)	7.76 (6.41)
TSK (10-item version)		41.47 (8.79)	36.41 (9.51)	34.99 (9.65)	35.27 (10.32)	34.11 (9.41)
RMDQ (23-item version)	**Control (*n*=121)**	11.36 (5.67)	11.44 (5.81)	10.01 (6.25)	8.99 (5.96)	9.02 (6.87)
TSK (10-item version)		41.25 (8.22)	39.96 (9.62)	39.17 (9.47)	37.55 (8.91)	38.58 (9.19)

7.8.3 Step 3.2: Combine both LGMs from Step 3.1 and model hypothesised relationships between the intervention and both LGMs

At this step the LGMs of the mediator and outcome were combined into a single model so that the relationships between both LGMs and intervention allocation could be investigated (Cheong et al 2003). This is described as a parallel process model, where the observed trajectories of the two measured variables are viewed as two separate processes that occur over the same period of

time (Cheong et al 2003). Including treatment group allocation as a co-variate also enables examination of the differences between scores in the intervention and control groups. In the Cheong paper, differences between the intervention and control groups were investigated using a two-group model, to see whether the factor loadings were appropriate to be used across both treatment groups (i.e. the factor scores used provided good fit to the data). However, as the overall mediation LGM would include treatment allocation as a co-variate to account for the different trajectories of the two groups, it was felt that this step was not necessary and that simply plotting the trajectories and choosing factor scores that best represented an overall trajectory for both groups in terms of model fit would be sufficient.

7.8.4 Step 3.3: Full mediation model and estimates of mediated effect (with bootstrapped CIs)

Cheong et al (2003) describe this step as using the parallel process model to estimate the mediating effect using the product of coefficients approach. As this approach, in conjunction with bias-corrected bootstrapped confidence intervals (CIs), was described as the most appropriate method of estimating mediation effects in Chapter 3, this was the approach adopted in the present analysis. Briefly, this step involves investigating a model which includes relationships between the intervention allocation and mediator variable (*a* path described in Figure 1.1 of Chapter 1) and between the mediator and outcome variable (*b* path). The two coefficients from each of these pathways are multiplied together to give a product of coefficients, or the mediated effect (*ab*). Including bootstrapped CIs helps to account for the often non-normal distribution of the mediated effect (Hayes 2009). In this analysis, 1,000 bias-corrected bootstrapped samples were used to estimate 95% CIs.

7.9 Results I: Non-response analysis and Imputation

Table 7.3 below shows how many people completed the key variables of interest at each time point. Overall, follow-up rates were good in this study, with less than 30 participants being lost to follow-up by 24 months in each group. The numbers of responders and non-responders was similar in both treatment and control groups with a maximum difference of three cases between responders and non-responders over the five time points. There were differences in baseline levels of pain intensity, fear-avoidance beliefs, disability and pain duration between responders and non-responders in the two treatment groups. However, these differences were not consistent, with the fear-avoidance beliefs scores of those in the treatment group being higher at baseline for responders to the earlier follow-up points and lower at baseline for those responding to later follow-up points. For disability, responders in the treatment group at two-month and six-month follow-up were more disabled at baseline and those responding at 12 and 24 months in the treatment group were less disabled compared to non-responders. For pain duration (days with pain), responders in the treatment group had fewer days of pain than non-responders at baseline for all time points except six-month follow-up. Overall, there does not seem to be a clear pattern of characteristics of non-responders in either the treatment or control group. However, the differences seen, while small and inconsistent, could potentially bias the results of the mediation analyses.

Table 7.3 Baseline characteristics (Means and SDs) of responders versus non-responders at two-, six-, 12- and 24-month follow-up

Variable		Baseline		Baseline values at two-month follow-up				Baseline values at six-month follow-up				Baseline values at 12-month follow-up				Baseline values at 24-month follow-up			
		Treatment *n*=119	Control *n*=121	Treatment		Control		Treatment		Control		Treatment		Control		Treatment		Control	
				R *n*=110	N *n*=9	R *n*=120	N *n*=1	R *n*=110	N *n*=9	R *n*=110	N *n*=11	R *n*=99	N *n*=20	R *n*=98	N *n*=23	R *n*=94	N *n*=25	R *n*=93	N *n*=28
Age		Mean 49.67 (SD 9.02)	Mean 49.82 (SD 9.77)	49.92 (8.85)	46.67 (11.03)	49.92 (9.75)	38.00 (0.00)	49.88 (8.87)	47.11 (11.01)	49.77 (9.87)	50.27 (9.14)	50.18 (8.78)	47.15 (9.98)	50.19 (9.71)	48.22 (10.08)	50.40 (8.92)	46.92 (9.05)	50.87 (9.02)	46.32 (11.43)
Sex		64.7% female	60.3% female	33.6% male	55.6% male	40% male	1 male	33.6% male	55.6% male	36.4% male	72.7% male	33.3% male	45.0% male	37.8% male	47.8% male	33% male	44% male	38.7% male	42.9% male
Education		Median 6.00 (range 3-8)	Median 6.00 (range 3-8)	7.00 (9.00)	4.00 (5.00)	6.50 (6.00)	5.00 (0.00)	6.00 (5.00)	7.00 (9.00)	6.50 (6)	6.00 (6)	7.00 (5.00)	6.00 (9.00)	7.00 (6)	6.00 (5)	7.00 (5.00)	5.00 (9.00)	7.00 (6)	6.00 (6)
Pain Intensity	Average last three months	Mean 5.71 (SD 1.84)	Mean 5.83 (SD 1.82)	4.90 (1.96)	3.56 (2.40)	5.30 (1.93)	0.00 (0.00)	4.19 (1.96)	3.44 (2.83)	4.72 (2.21)	2.82 (2.60)	4.04 (2.28)	4.15 (3.45)	4.72 (2.14)	4.39 (3.49)	4.26 (2.07)	3.60 (3.11)	4.61 (2.49)	4.18 (3.27)
	Today	Mean 3.45 (SD 2.49)	Mean 3.45 (SD 2.57)	2.87 (2.35)	2.33 (2.69)	3.14 (2.51)	2.00 (0.00)	2.76 (2.49)	1.56 (3.32)	3.05 (2.63)	3.00 (2.37)	2.49 (2.67)	3.55 (3.61)	2.73 (2.54)	3.43 (3.22)	2.34 (2.46)	2.52 (3.08)	2.77 (2.86)	3.29 (3.18)
Pain Duration (days)		Mean 110.50 (SD 60.72)	Mean 97.39 (SD 62.34)	113.46 (64.25)	122.60 (69.01)	101.80 (63.50)	144.23 (0.00)	94.99 (74.98)	32.42 (94.61)	88.67 (68.86)	98.79 (118.51)	93.29 (72.94)	140.64 (329.64)	91.52 (71.48)	67.48 (121.90)	82.56 (75.99)	129.07 (298.74)	73.45 (72.07)	68.57 (113.67)

Variable	Baseline		Baseline values at two-month follow-up				Baseline values at six-month follow-up				Baseline values at 12-month follow-up				Baseline values at 24-month follow-up			
			Treatment		Control		Treatment		Control		Treatment		Control		Treatment		Control	
	Treatment *n*=119	Control *n*=121	R *n*=110	N *n*=9	R *n*=120	N *n*=1	R *n*=110	N *n*=9	R *n*=110	N *n*=11	R *n*=99	N *n*=20	R *n*=98	N *n*=23	R *n*=94	N *n*=25	R *n*=93	N *n*=28
Fear-Avoidance beliefs (10-item TSK)	Mean 41.47 (SD 8.79)	Mean 41.25 (SD 8.22)	36.38 (9.27)	37.37 (14.58)	39.89 (9.66)	47.23 (0.00)	35.13 (9.86)	30.74 (9.30)	39.21 (9.66)	37.60 (9.07)	34.28 (9.95)	42.40 (23.94)	37.39 (9.46)	36.07 (9.37)	34.21 (9.70)	37.60 (10.80)	38.44 (9.89)	37.81 (7.76)
Disability (23-item RMDQ)	Mean 12.10 (SD 5.49)	Mean 11.36 (SD 5.67)	10.24 (6.32)	7.60 (8.11)	11.47 (5.83)	6.99 (0.00)	9.23 (6.56)	2.90 (10.42)	10.13 (6.39)	10.87 (6.48)	8.43 (6.96)	13.09 (22.64)	9.10 (6.34)	10.53 (7.12)	8.09 (6.46)	9.69 (21.49)	9.14 (7.16)	10.05 (10.41)

R=Responder; N=Non-responder

EM imputation was therefore carried out cross-sectionally on the baseline and each follow-up dataset. Little's MCAR test was non-significant for the baseline, six-month, 12-month and 24-month datasets but produced a statistically significant result in the two-month dataset.

7.10 Results II: Preliminary analysis

7.10.1 Preliminary analysis: Descriptive statistics

The mean (SD) scores for the potential mediator and outcome variables in the two intervention groups are given in Table 7.2. As expected, the biggest mean change in the intervention group was between baseline and the first follow-up point of two months for both the potential mediator and the outcome variable. The reduction in fear-avoidance beliefs and disability over the two-month intervention period was smaller for the control group receiving usual care compared to the intervention group. The numbers lost to follow-up were similar in both the treatment and control groups (see Table 7.3), suggesting that the control group were as engaged in their treatment as the treatment group. The skewness and kurtosis values (explained in Chapter 5) were below 1.0 for all variables except for kurtosis at two months (-1.00), six months (-1.15) and 12 months (-1.21) for RMDQ score in the treatment group, and kurtosis at six months (-1.01) and 24 months (-1.23) for RMDQ score in the control group. Overall, these scores suggest that the variables were normally distributed, which is an assumption underlying LGM with maximum likelihood (ML) estimation.

7.10.2 Psychometric properties of the TSK (10-item) and RMDQ (23-item)

As described in Chapter 5, the 24-item RMDQ has been found to have good psychometric properties overall in LBP populations (e.g. Roland & Morris 1983; Roland & Fairbank 2000). The 23-item version of the RMDQ used in Von Korff et al's (2005) study has received less attention in

terms of its psychometric properties, although studies do exist which show it to have good test-retest reliability and internal consistency in a primary care LBP population (Dunn et al 2003; Dunn & Cherkin 2007). In the present study, Cronbach's alpha showed that the 23-item RMDQ had a good internal consistency (0.88) in this population (Table 7.4).

There has been much debate over the factor structure of the original version of the TSK (described fully in Chapter 5). The version of the TSK used in the present study was modified by the original study authors (Von Korff et al 2005) and was assumed to have a single factor structure. In the present analysis this version of the tool was found to have an internal consistency value of 0.66 (Table 7.4). Cronbach's alpha values of 0.70 or above are seen as showing adequate internal consistency. This means that this measure may not be reliable, as suggested by studies on the factor structure of other versions of this tool. Reports of internal consistency for other versions of the tool have been inconsistent, with values of between 0.73 (Lamé et al 2008) and 0.84 (French et al 2007) found in samples of chronic LBP patients using the original TSK, but these values will be influenced by the larger number of items in the original measure. However, as it is the only available measure of fear-avoidance beliefs in the dataset, it will still be used in the present analysis and kept in the single-factor structure used in the original study.

Table 7.4. Measures of internal consistency

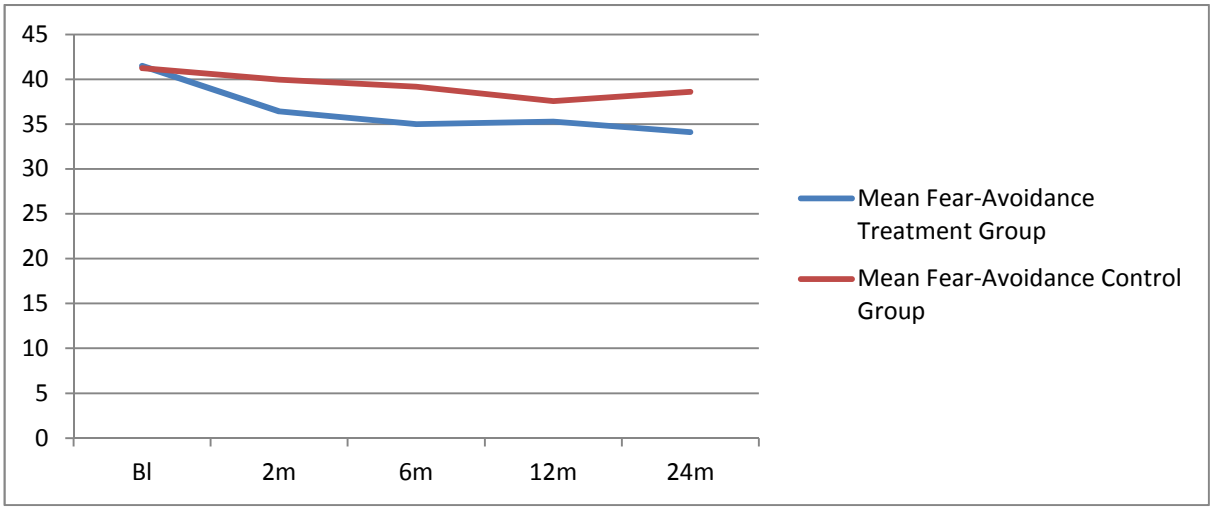
Measure	**No. of Cases Analysis Based on**	**Cronbach's Alpha Value**
TSK (10-item version)	240	0.66
RMDQ (23-item version)	240	0.88

7.11. Results III: Mediation Analysis: Latent Growth Modelling

7.11.1 Observed trajectories of mediator and outcome variables (Step 3.1)

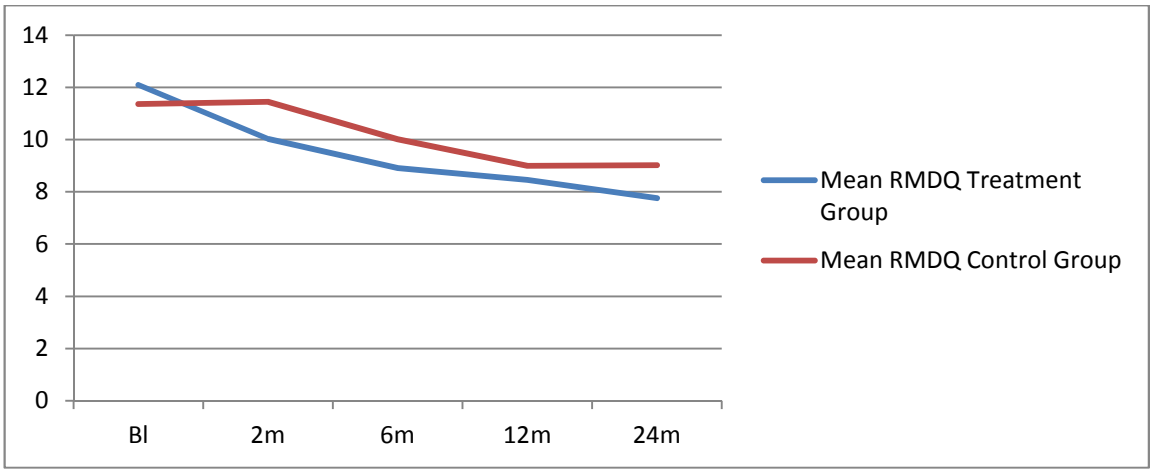
The observed trajectories of mean fear-avoidance beliefs scores for the treatment and control groups based on all measured time points are presented in Figure 7.3a. The results show that, due to successful randomisation, the scores were very similar at baseline (less than one-point difference between treatment and control) but that the scores in the treatment group decreased substantially by two-month follow-up compared to the control group. This difference in scores was smaller at 12-month follow-up but larger again by 24-month follow-up. However, it is clear from the graph that neither observed trajectory followed a linear pattern.

Figure 7.3a Course of 10-item TSK scores (mediator variable) for intervention and control groups over two years follow-up



The observed trajectories of the outcome measure in the treatment and control groups were then examined (Figure 7.3b). The treatment group showed a linear decrease in disability scores compared to the control group, which showed no change in disability over the first two months and also no change between 12-month and 24-month follow-up. The mean RMDQ scores at baseline showed a small difference (0.7 point) despite randomisation.

Figure 7.3b. Course of 23-item RMDQ scores (outcome variable) for intervention and control groups over two years follow-up



In LGM, these observed trajectories were incorporated by the slope factor of the model using the factor loadings, or factor scores (see Figure 7.2). The intervals at which the measures were taken also need to be taken into account when estimating these scores. Figure 7.2 shows potential factor loadings if the measures were taken at equal time intervals and if the scores represented a positive linear trend. However, the Back In Action trial does not have equally spaced measurement intervals and the observed course of scores for the two intervention groups did not follow a linear trend. To take account of this, the factor scores of the slope factors for the mediator and outcome were estimated based on the observed trajectories, taking into account the unequal measurement points (see Figures 7.4a and 7.4b).

The factor scores were based on the observed trajectory for both the treatment and control groups combined. This will result in a better-fitting model than if just the treatment group scores were used, as the groups will be compared using the co-variate of treatment group allocation in the full mediation model. The factor scores for both the mediator and outcome trajectories appeared to provide an adequate fit to the data, as indicated by the goodness-of-fit statistics

(Tables 7.5a and 7.5b). CMIN was statistically significant, indicating poor fit. The RMSEA value also indicated poor fit. However, the remaining indices suggested good fit, so overall it was agreed that the model provided adequate fit to the data. Judgement of model fit statistics is subjective as no one index is superior to any other, so a decision was made based on the results of all the indices collectively.

Figure 7.4a. LGM model for fear-avoidance beliefs (10-item TSK)

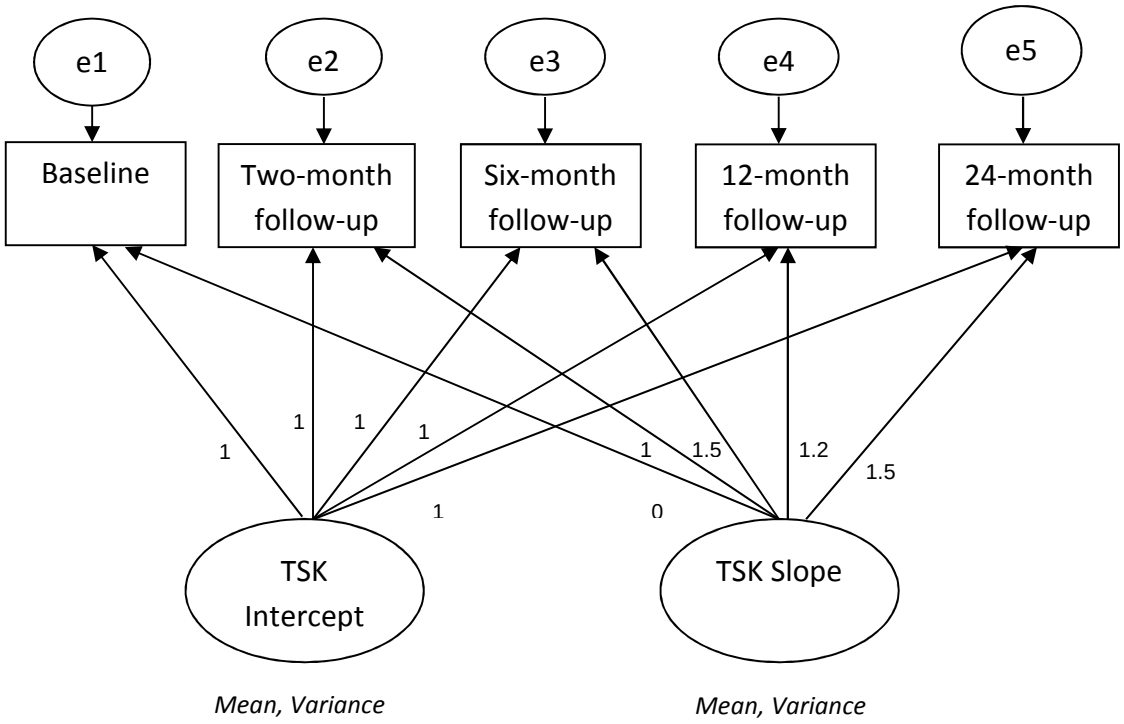


Table 7.5a. Model fit statistics for growth curve model of 10-item TSK

Model Index	**Current Model**	**Good Model Fit**
CMIN (Chi Square)	62.53	Non-significant result
DF	14	
P	0.00	
CMIN/DF	4.47	Between 2-5
CFI	0.95	>0.95
RMSEA	0.12 (90% CI 0.09 to 0.15), *p*<0.01	<0.08
SRMR	0.05	<0.08

Figure 7.4b. LGM model for disability (23-item RMDQ)

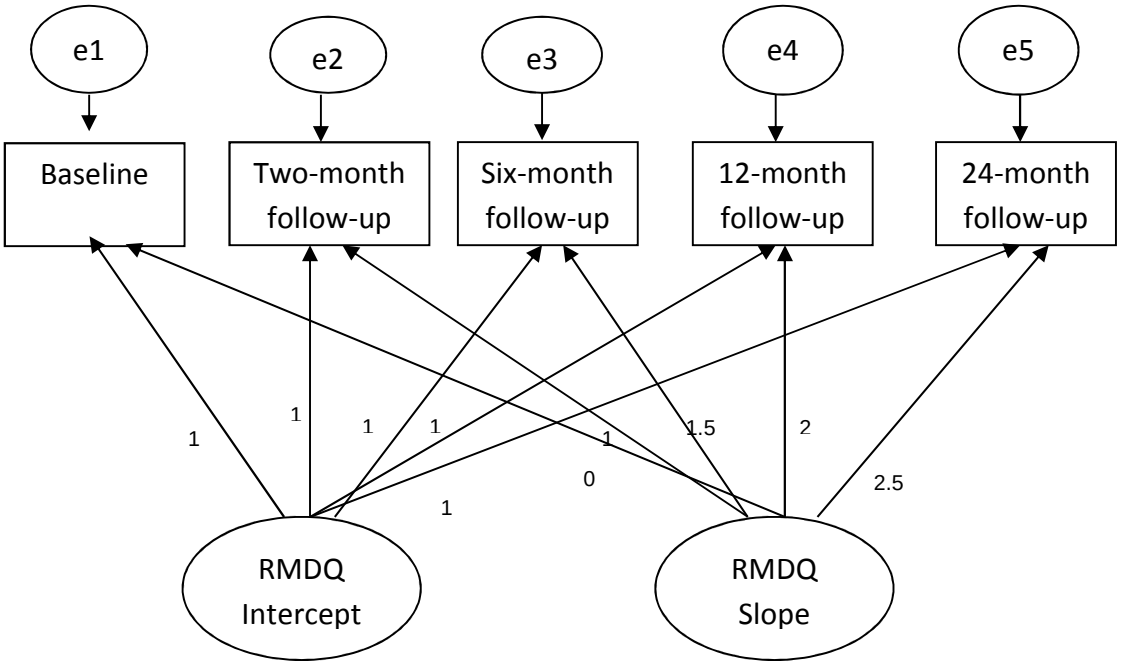


Table 7.5b. Model fit statistics for growth curve model of 23-item RMDQ

Model Index	**Current Model**	**Good Model Fit**
CMIN (Chi Square)	49.95	Non-significant result
DF	14	
P	0.00	
CMIN/DF	3.57	Between 2-5
CFI	0.96	>0.95
RMSEA	0.10 (90% CI 0.07 to 0.14), *p*<0.01	<0.08
SRMR	0.04	<0.08

7.11.2 Latent Growth Modelling: Parallel Process Models (Step 3.2)

The next step was to run the models together to test for co-variances between each of the Intercept and Slope factors and explore how the potential mediator and outcome variables relate to each other, before adding the intervention variable to complete the model. Crucially, it is important to test whether change in the potential mediator is related to change in the outcome variable, as this is a necessary preliminary step in mediation analysis. At this stage, while this model examines how change in the potential mediator and outcome variables is related, it does

not look at differences between the intervention and control groups. A simplified model is shown in Figure 7.5 below. This figure does not include the measurement part of the model or error terms, but these are the same as those presented in Figures 7.4a and 7.4b. The model fit statistics in Table 7.6 indicated that this model provided only an average fit to the data overall, so further investigation was needed to see if model fit could be improved. The standardised residuals were all within the ±1.96 threshold for the original model, suggesting that all were within the acceptable range. Modification indices were therefore examined, and it was found that allowing co-variation between the error terms for the TSK and RMDQ 12-month follow-up scores and between the error terms for the TSK and RMDQ two-month follow-up scores slightly improved model fit (see Table 7.6). As it is likely that each of the measurement points are related to each other, and that this may be reflected in related error variance (Campbell et al 2013b), it was decided that allowing these co-variances made theoretical sense. Further modifications did not appear to be theoretically plausible and therefore no further changes to the model were made.

Figure 7.5 Simplified LGM model to show relationships between mediator and outcome variables

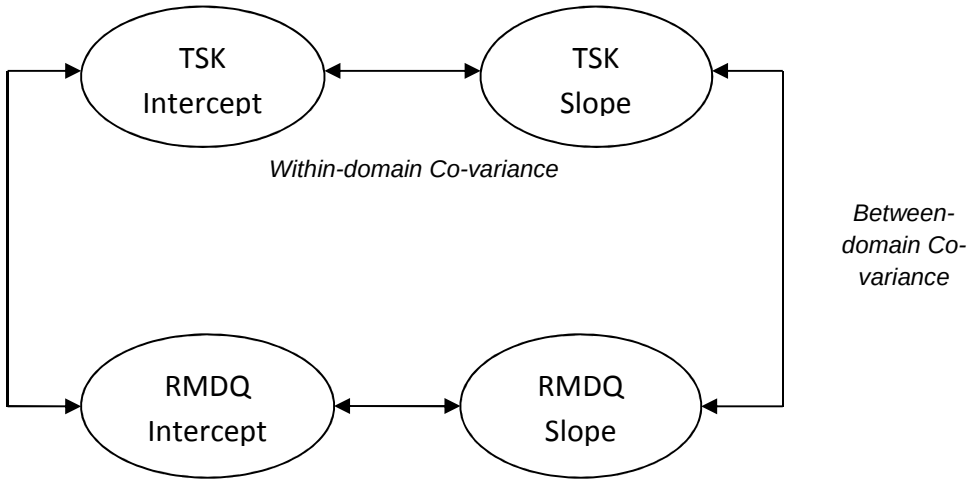


Table 7.6 Model fit statistics for parallel process model without treatment group allocation as a co-variate

Model Index	Original Model	Modified Model	Good Model Fit
CMIN* (Chi Square)	217.71	154.88	Non-significant result
DF	51	49	
P	0.00	0.00	
CMIN/DF	4.27	3.16	Between 2-5
CFI	0.91	0.94	>0.95
RMSEA	0.12 (90% CI 0.10 to 0.13), p<0.01	0.10 (90% CI 0.08 to 0.11), p<0.01	<0.08
SRMR	0.05	0.05	<0.08

Table 7.7 gives the results of the means, co-variances and variances between the intercept and slope of each variable. The mean change in both TSK and RMDQ indicated a slight, statistically significant decrease over time in both variables. The within-domain co-variance, relating to the intercept and slope of the same construct (Byrne 2010), showed a positive estimate between TSK intercept and slope and a negative estimate between RMDQ intercept and slope. This suggested that people who scored highly on the TSK at baseline had a higher rate of decrease in scores than those who scored lower, while for the RMDQ, people who scored higher at baseline had a lower rate of decrease in scores. However, both of these estimates were very small and did not reach statistical significance. The between-domain co-variances gave much higher, statistically significant values. The high value between the TSK and RMDQ slope factors suggested that as TSK scores change the RMDQ scores also change. The high value between the two intercepts suggested that people who scored highly on the TSK are likely to also score highly on the RMDQ. Finally, the statistically significant variances for each latent factor indicated strong individual variation amongst participants, especially in the initial scores. This is likely to be due to the fact that primary care patients with chronic LBP are a very heterogeneous population who are likely to have variable levels of disability and fear-avoidance beliefs. To test for an intervention effect, the next step is to add treatment group allocation as a co-variate in the analysis.

Table 7.7 Means, co-variances, correlations and variances between mediator and outcome variables

	Variables	**Estimate**	**Standard Error**	***p*-value**
Means	TSK Intercept (*baseline value*)	41.14	0.56	0.00
	TSK Slope (*change over all time points*)	-3.06	0.32	0.00
	RMDQ Intercept	11.81	0.36	0.00
	RMDQ Slope	-1.38	0.15	0.00
Co-variances	TSK Intercept <-> TSK Slope	β= 1.06 b= 0.05	2.71	0.49
	TSK Intercept <-> RMDQ Intercept	β= 20.79 b= 0.60	3.10	0.00
	TSK Slope <-> RMDQ Slope	β= 3.64 b= 0.86	0.68	0.00
	RMDQ Intercept <-> RMDQ Slope	β= -0.31 b= -0.04	0.79	0.69
Variances	TSK Intercept	52.05	6.69	0.00
	TSK Slope	7.93	2.30	0.00
	RMDQ Intercept	23.06	2.81	0.00
	RMDQ Slope	2.24	0.50	0.00

β = *Unstandardised estimate*
b = *Standardised estimate*

Figure 7.6 shows the LGM for the mediator with treatment group allocation as a co-variate. This model is identical to the previous model but provides additional information about whether the initial status and change in the mediator and outcome variables differ as a result of treatment group allocation. The model provided a very similar fit to the data as the model without the treatment group allocation variable (Table 7.8). The standardised residual co-variances were therefore examined, which suggested that a large number of parameters should be unconstrained. However, too many co-variances exceeded the ±1.96 threshold for changes to be feasible. Modification indices were therefore examined, which indicated that allowing co-variation between the error terms for TSK and RMDQ scores at 12-month follow-up, TSK score at 12-month follow-up and RMDQ score at two-month follow-up, and TSK and RMDQ scores at two-month follow-up would result in a better fitting model (see Table 7.8). The co-variation between TSK score at 12-month follow-up and RMDQ score at two-month follow-up is less intuitive than

co-variances between the variables at the same time points, but could reflect a relationship between early change in disability and later change in fear-avoidance beliefs.

In the present analysis the values of interest are the regression coefficients between treatment allocation and the slopes and intercepts for the mediator and outcome (Byrne 2010). Table 7.9 shows that intervention group allocation significantly predicted the rate of change for both the TSK and RMDQ measures, but not the initial status, which is expected given random allocation of participants to intervention groups. As the intervention variable was coded 0 for the control group and 1 for the intervention group, the slope values for TSK and RMDQ suggest that change in the TSK and RMDQ measures was substantially higher in the intervention group, particularly for the change in TSK. This is indicated by the estimates of -0.26 and -0.65 for TSK and RMDQ scores respectively (see Table 7.9), showing how they have decreased over time in the intervention group compared to the control group.

Figure 7.6 Simplified LGM model to show relationships between mediator and outcome variables, with treatment group allocation as a co-variate

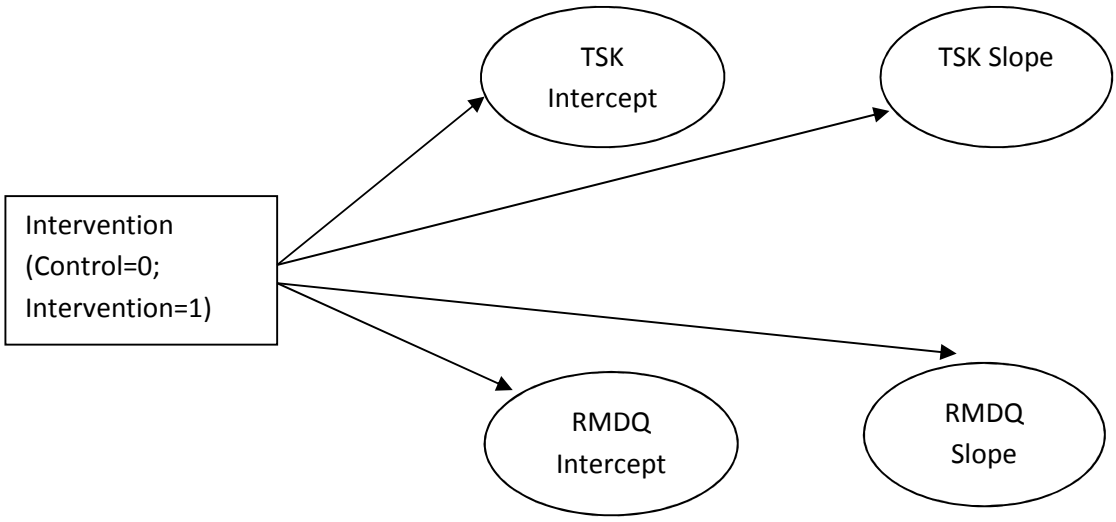


Table 7.8 Model fit statistics for parallel process model with treatment group allocation co-variate

Model Index	**Original Model**	**Modified Model**	**Good Model Fit**
CMIN (Chi Square)	237.21	172.17	Non-significant result
DF	57	54	
P	0.00	0.00	
CMIN/DF	4.16	3.19	Between 2-5
CFI	0.91	0.94	>0.95
RMSEA	0.12 (90% CI 0.10 to 0.13), *p*<0.01	0.10 (90% CI 0.08 to 0.11), *p*<0.01	<0.08
SRMR	0.05	0.04	<0.08

Table 7.9 Mean values for mediator and outcome, related to treatment group allocation

	Variables	**Estimate**	**Standard Error**	**p-value**
Regression Coefficients	TSK Intercept	β = 0.10 b= 0.01	1.12	0.93
	TSK Slope	β= -0.26 b= -0.45	0.61	0.00
	RMDQ Intercept	β=0 .21 b= 0.02	0.72	0.77
	RMDQ Slope	β= -0.65 b= -0.22	0.29	0.02

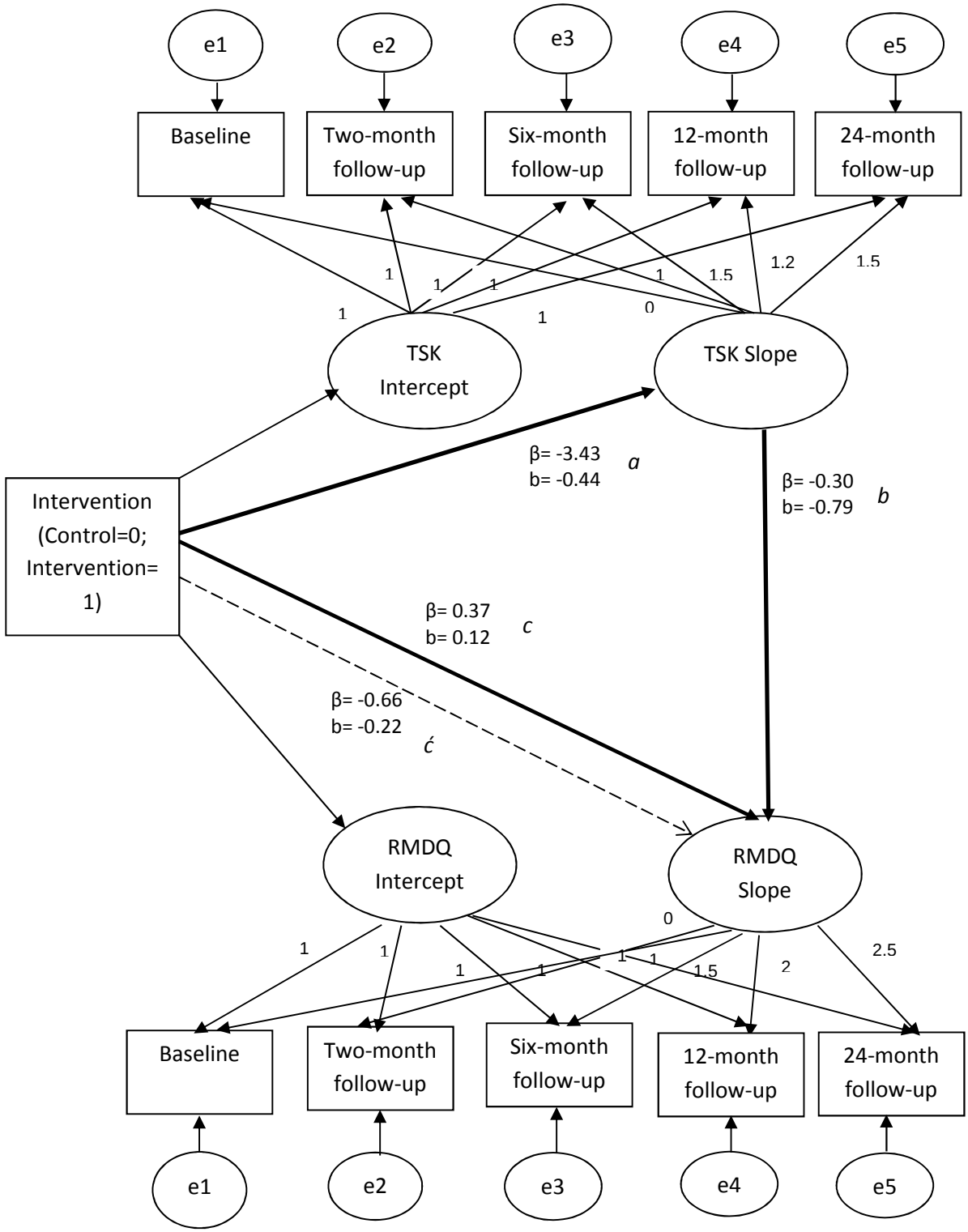
β = *Unstandardised estimate*
b = *Standardised estimate*

7.11.3 Latent growth modelling: Mediation analysis (Step 3.3)

Finally, the full mediation model was tested by adding a path between the mediator and outcome variables. The bold lines on Figure 7.7 below highlight the mediating pathway. Each of the beta coefficients shown on this Figure represent the direct effect of each variable on the other, as indicated by the direction of the arrow. However, it must be acknowledged that LGM, like SEM, is a correlational technique and therefore it cannot be established whether the effect is in the hypothesised direction. The standardised coefficients show that the intervention is associated with a decrease in the TSK slope (-0.44) and that this decrease is associated with a decrease in the RMDQ slope (-0.79). There is a small direct effect (0.12) on RMDQ slope but this is not in the same direction (positive) as the other coefficients. Table 7.10a shows that the mediated effect (product of coefficients; in this case, -0.44 * -0.79 = -0.35 (95% CI -0.47 to -0.24)) was statistically

significant, indicating that differences between intervention and control in change in the observed trajectory of the fear-avoidance beliefs variable mediated some of the change in the observed trajectory of the outcome measure. However, none of the model fit statistics for this final model suggest that it provided a good fit to the data (see Table 7.10b) and modifying the model using the standardised residual co-variances and modification indices did not help to improve fit. This means that the model as currently hypothesised does not fully explain how the treatment worked for this particular patient population, as there is key information missing. The study hypothesis, that allocation to the activation intervention arm was associated with a reduction in fear-avoidance beliefs and that this reduction was associated with a reduction in disability outcome, was therefore partially supported, but suggests that other mediating factors must be considered.

Figure 7.7 Full mediation model: Change in fear-avoidance beliefs (10-item TSK) as a mediator between treatment allocation and change in disability outcome (23-item RMDQ)



β = *Unstandardised estimate*
b = *Standardised estimate*
NB: Total effects are taken from Table 7.10a

Table 7.10a Mediating effect of change in fear-avoidance beliefs (10-item TSK) on the relationship between treatment allocation and change in disability (23-item RMDQ)

	Effect	Model	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ (23-item)$^{\Delta}$	Total (*ć*)	-0.22 (-0.40 to -0.02)	-0.66 (-1.20 to -0.08)
	Direct (*c*)	-0.11 (-0.11 to 0.33)	-0.31 (-0.31 to 1.01)
	Indirect (*ab*)	-0.35 (-0.48 to -0.23)	-1.03 (-1.49 to -0.67)

$^{\Delta}$=*residualised change*

Table 7.10b Model fit statistics for full mediation model

Model Index	**Full Mediation Model**	**Good Model Fit**
CMIN (Chi Square)	1095.17	Non-significant result
DF	63	
P	0.00	
CMIN/DF	17.38	Between 2-5
CFI	0.46	>0.95
RMSEA	0.26 (90% CI 0.25 to 0.28), *p*<0.01	<0.08
SRMR	0.36	<0.08

7.12 Discussion

7.12.1 Summary

The focus of this chapter was to test an analytical approach, recommended by the literature, in a dataset from an RCT and also to test the study's hypothesis that change in fear-avoidance beliefs helped to explain the reduction in disability scores seen in the intervention arm of the trial (i.e. that change in fear-avoidance beliefs was a mediator of the relationship between treatment allocation and disability improvement).

The technique of LGM used to carry out this analysis proved a useful, if complex, technique for mediation analysis. It allowed for the conduct of the preliminary steps outlined in the previous chapters, but unlike the previous analyses it allowed a number of time points to be included

simultaneously in the model. As one of the requirements for future trials set out in Chapter 3 was an increased number of time points to be included in studies, this technique is likely to prove useful in future mediation analysis studies.

The results of this analysis showed that there were associations between the intervention and change in fear-avoidance beliefs, and between change in fear-avoidance beliefs and change in disability. There was a significant mediating effect of change in fear-avoidance beliefs on the relationship between treatment allocation and change in disability. This suggests that targeting fear-avoidance beliefs in an intervention for chronic LBP patients would be beneficial. However, the final mediation model provided a poor fit to the data, suggesting that key information is likely to be missing from the hypothesised model and that other mediating factors should be considered.

7.12.2 Comparison with previous findings

This analysis adds to a currently small evidence base for treatment mediators of LBP interventions (see Chapter 4). It is the first study, to the author's knowledge, that has used LGM for analysing mediators in a LBP intervention study. Having a specific focus during an intervention, in this case on addressing fear-avoidance beliefs, may help make the study more suitable for mediation analysis, provided the variable of interest is well measured and measured often, and may also be more likely to demonstrate a stronger mediating effect than an intervention which has a less specific focus on a particular construct.

This study also adds to the small but growing number of studies that have used LGM as a technique for examining mediation with RCT data. Guidelines from previous LGM studies in other areas of health psychology were followed (Cheong et al 2003; Roesch et al 2009). However, it is difficult to compare the mediating effects found in these studies with those found in the present

study as different populations were examined and the intervention length differed (for example Cheong et al's study included booster sessions to help participants retain knowledge of the intervention and Roesch et al's intervention lasted 12 months), which could have an effect on the strength of any mediating effects found. A recent study also claimed to use Cheong et al's technique to carry out their mediation analysis on a RCT data from an intervention to prevent child behavioural problems (Piehler et al 2014), but while this study utilised LGM they obtained their mediating effect using the Baron & Kenny approach rather than using the product of coefficients.

7.12.3 LGM as a mediation analysis technique

Using LGM to test for mediating effects has several advantages over using SEM, specifically the advantage of being able to study change across more than two time points. This gives it the potential to address a key problem with current mediation studies, temporality (see Section 7.12.4) which is not currently adequately addressed. However, the use of LGM alone, as this study shows, cannot address temporality without including an adequate number of time points that will capture change when it occurs. Also, unlike SEM it cannot include latent variables in the analysis. This does have the disadvantage of assuming no measurement error in the tools used to assess the variables of interest, an assumption which is difficult to make when assessing psychological factors (See Chapter 3, Section 3.2.5). The specific measures used in this study, particularly the measure used for the potential mediator (10-item TSK), were not well validated in this population (see Section 7.10.2), and the low internal consistency value found for the modified TSK (0.66) suggests that the measure may not be reliable. The use of such a different technique in the present analysis also precludes comparison with the SEM analysis carried out in Chapter 6, which included latent variables and fewer time points. However, the purpose of this analysis was to use the method most suited to answering the study hypothesis as to how the intervention worked,

and also to improve the validity of the mediation analysis by using optimal methods, rather than compare results across different studies. The study hypothesis and design of this study is different to that used in the STarT Back trial (Chapter 6) and therefore required different analysis methods.

One problem encountered when conducting this analysis was the lack of studies that had applied this technique to mediation analyses of intervention studies previously, and it was the first time that this approach had been used in this topic area. This meant that there was little guidance around how to carry out the analysis (Cheong et al 2003; Roesch et al 2009; Byrne 2010) so time had to be taken to adequately develop and apply the model in order to produce robust results. Some of the decisions made in this analysis, such as which factor score values to assign to the slope factors, are based on personal judgement on what best seemed to fit the data. Whilst it was felt that such decisions were justifiable and sensitivity analyses to compare other models was carried out (see Appendix 7.2) it is possible that these models may not be generalisable to other intervention studies.

7.12.4 Potential limitations

Similar to the STarT Back trial the intervention in this trial was complex, involving several sessions with a multidisciplinary team and several different components. While several of these sessions did focus on reducing fear-avoidance beliefs, much of the intervention actually focused on goal-setting and on plans to achieve those goals. These are likely to be highly variable depending on the individual patient, and could depend on factors such as self-efficacy or internal control, or the patient's relationship with their therapist. These factors could also potentially be treatment mediators along with fear-avoidance beliefs but they were not considered as mediating factors in the original study and no measures of them were provided, so they could not be tested.

It could also be that the population included in the intervention did not score particularly highly on the fear-avoidance beliefs measure used. However, little information is available on an acceptable cut-off for a 'high' TSK score, and no information is available on the 10-item measure used in this study. Previous studies using the Swedish and Dutch versions of the 17-item TSK have recommended a cut-off of 37 for 'high' fear (Vlaeyen et al 1995a; Vlaeyen et al 1995b; Lundberg et al 2006). In this study, the trial authors scored the 10-item TSK in a way that made it comparable with the 17-item version. If these cut-offs are applied to the present study, it would appear that both the treatment and control groups have high fear at baseline (mean score over 37) and while in the treatment group this score decreases (below 37 at each follow-up point), in the control group the scores at each follow-up point remain above this cut-off.

The tools used to measure the variables of interest in this study, the RMDQ (functioning) and TSK (fear-avoidance beliefs), were modified versions of the original tools. The RMDQ provided an adequate value of internal consistency and other aspects of its psychometric properties are supported by evidence from other sources (see Section 7.10.2) but the modified version of the TSK used in this study did not. This finding is supported by literature on the factor structure of other versions of the TSK. However, internal consistency is only one property that can affect a measure's performance; responsiveness is another important measurement property when looking at mediation, as we are looking at change over time, and this property warrants further investigation for the TSK. This study did demonstrate that significant change occurred across the time points used in the study, but this is based on how well the measures assessed change. In future studies therefore, the measurement properties of the tools used in mediation analysis ideally should be assessed prior to their inclusion in the study as they may have an impact on the results of the analysis.

One of the key reasons given in the literature for using LGM for mediation analysis is that it allows the inclusion of several time points. This helps to better assess temporality, or when the variables change in relation to each other. Despite its increased complexity compared to the analyses presented in Chapters 5 and 6, this analysis still cannot address the issue of temporality, as although the model hypothesised that the intervention led to a change in fear-avoidance beliefs which in turn led to a change in disability, it could perhaps be that the change in disability occurred before or concomitantly with the change in fear-avoidance beliefs. This was because while the Back In Action study collected data across five time points, it did not collect any data while the intervention was taking place. This limits the true assessment of treatment-mediated effects as it is likely that the follow-up scores show a more diluted effect of the intervention than at the time of participants receiving it. Figures 7.3a and 7.3b show that the change in both fear-avoidance beliefs and disability happens early on, but it is still not possible to see when these variables changed and crucially, *which* changed first. One way of potentially addressing this issue would have been to look at the intercept of the mediating variable and the slope of the outcome, rather than the slopes of both variables (to see whether baseline levels of fear-avoidance beliefs predict change in disability) (Selig & Preacher 2009), but that would have been illogical in this analysis as the hypothesis was around change in the mediator as well as the outcome. Looking at change in the mediator at an earlier stage than change in the outcome is another option, but again not possible in this analysis as the first follow-up point was two months post-treatment, at which any change that was expected to happen in both the mediator and outcome variables would have already occurred. However, such analysis is possible using LGM provided that there are enough time points to allow examination of early and later change. This again stresses the need for assessment points during treatment.

Imputation was carried out on this dataset in order to obtain statistics required for mediation (bias-corrected bootstrapped CIs) and model fit (SRMR). Overall there was only a small amount of missing data throughout the follow-up periods, but the test for whether or not data was missing at random indicated that at two-month follow-up, data was not MCAR but MAR or possibly NMAR. However, the dropout at the two-month follow-up point was very small (*n*=10 (8% of total sample)), so this follow-up point was still included in the analysis as it was felt that the risk of bias due to dropout at this follow-up point would be low.

The issue of model fit in SEM and LGM is challenging. No consensus exists over which indices are most useful to focus on, and different authors give different cut-off points for good fit which adds to the uncertainty and subjectivity when attempting to interpret these values. If a model does not fit well and improvements cannot be made to it to improve the fit, it is unclear whether the answer is to reject the hypothesised model outright (i.e. change in fear-avoidance beliefs is not a mediator of the effect of treatment allocation on change in disability); to accept that the model does not provide the complete picture as to what factors mediate the effect of treatment allocation on outcome in this study (i.e. fear-avoidance beliefs is a mediator but is not the only factor in the pathway - a highly likely conclusion in any mediation study); or whether there are other issues that have impacted on model fit such as methodological errors (e.g. timing of measurements, or inadequate internal consistency of the potential mediator). Byrne (2005) states that a lack of model fit implies implausible relationships between the variables, but it is still unclear whether this means there is information missing from the model or whether the model is simply wrong. It is also unclear what level of importance should be given to the fit indices – equivalent models may fit just as well (Tomarken & Waller 2003), and even if a model fits well it may still be misspecified (Tomarken & Waller 2003; Hayduk 2014). Many authors have expressed caution over making modifications to models that do not fit well (e.g. Tomarken & Waller 2003;

Byrne 2010), as additional paths may mean the model cannot be generalised to other samples (Tomarken & Waller 2003). In the present analysis, attempts were made to modify the models where necessary at each stage to improve fit, but these modifications were only made if it was felt that the modification made theoretical sense. The final full mediation model in the present study did not fit the data sufficiently, despite attempts to improve fit through the use of modification indices. I have interpreted this as suggesting that the model is incomplete, with key mediating variables missing from the model. This means that the hypothesis that allocation to the treatment arm would lead to a reduction in fear-avoidance beliefs and subsequently to an improvement in disability was partially supported, as while fear-avoidance beliefs is clearly not the sole variable on the mediating pathway it is still an important mediating factor.

7.12.5 Clinical Implications

Clinically, the findings from this study suggest that fear-avoidance beliefs play a role in the explanation of treatment outcome when a treatment specifically focuses on targeting this factor. Future treatments should perhaps include only patients who are found to be highly fear-avoidant, or tailor treatment to focus fear-avoidance belief reduction on patients who are most severe. However, this analysis represents only one study and more research is needed to support these findings, preferably research that accounts for the limitations discussed above. Few additional variables that could have played a key role in the success of this intervention (such as self-efficacy) were measured, and the role of other factors in this intervention are therefore unclear. Another concern is that the measure of fear-avoidance beliefs available in this dataset has not been validated previously, and its internal consistency in this dataset was poor. A stronger measure of fear-avoidance beliefs, to ensure more confidence that this construct is indeed a mediator of outcome, is required in future studies.

In order to improve the delivery of future interventions, it would also be helpful to see which aspect of the intervention led to a change in fear-avoidance beliefs. For example, if this was after the first session, when fear of pain was explicitly discussed, it could be that this session was the one that was important in changing fear-avoidance beliefs. Measurements taken during the treatment period would give much more information on key variables that the intervention aimed to change, so that future interventions can be tailored to only include sessions found to impact on factors of importance.

7.13 Conclusion

Using LGM to investigate mediating factors was found to be a valid and useful technique when investigating studies with multiple time points. This type of analysis will be useful in future studies of mediation to allow exploration of the temporal nature of variables of interest. The reduction of fear-avoidance beliefs through a multidisciplinary intervention appeared to be a mediating factor for the reduction of disability in chronic LBP patients. However, a key problem still not resolved by this study is that of temporality. While this study included multiple time points and showed that LGM can be used to assess all of these time points when testing for treatment mediation, none of the assessments were made during the intervention period. The next chapter in this thesis provides analysis of assessments made during the intervention period by means of a diary study, which will help answer the final thesis objective of exploring treatment processes.

Chapter 8: Using Diary Data to Examine the Importance of Measuring During Treatment for the Analysis of Mediating Factors: IMPACT Dataset

8.1 Introduction

This chapter builds on previous chapters which highlighted the main issues encountered when conducting mediation analysis and the different methods that can be used to try and counteract these problems. The present chapter examines a particular issue, temporality, which has been difficult to address so far. One of the main reasons for this difficulty is the lack of studies that have measured key mediators and outcomes during the treatment process. The analysis presented here includes constructs that were measured at weekly intervals during the participants' treatment period, as well as at pre-treatment, post-treatment and three-month follow-up. The purpose of this chapter is to examine whether temporality can potentially be established by the inclusion of measures during treatment, as stated in the literature, and also whether the inclusion of such measures will help improve future treatment mediation analyses.

8.2. Rationale for Study

Several authors have highlighted that a key design issue in studies of mediation of treatment effect is establishing temporality, or change over time. Many have recommended that a larger number of repeated assessments should be included in randomised controlled trials (RCTs) so we can establish not only efficacy, but also the processes of change (Kazdin 2007; Laurenceau et al 2007; Burns et al 2015). Assessing variables only at pre- and post-treatment, as is typically done in RCTs, cannot tell us anything about how factors change over time. In fact, it leads us to assume a linear relationship between potential mediators and outcome which is unlikely to be the case

(Laurenceau et al 2007). Examining individual trajectories of change over time may therefore be more meaningful in identifying mediators (Kraemer et al 2002; Laurenceau et al 2007).

A related and important aspect is *when* these assessments take place. It is important that the hypothesised timing of when mediators are likely to change is taken into account when deciding when to assess variables in the study (Laurenceau et al 2007; Preacher 2015), as the use of different measurement intervals are likely to expose or miss relationships between variables (Collins & Graham 2002) depending on the true underlying mediating process. For example, if measurements are taken too far apart in time or too early on, relationships might be missed as the effects may take some time to emerge. It is therefore very important to decide *a priori* when the best measurement time-point is for each variable of interest (Laurenceau et al 2007). Assessing variables at only two time points also raises issues around measurement error (Laurenceau et al 2007). While adding a mid-point assessment has been suggested as a way of improving this issue, this point is often arbitrary and not based on theory or research findings (Laurenceau et al 2007) meaning that mediating effects may still be missed.

However, it is also acknowledged that the decision of when to measure key variables is not easy to make currently, as there is little available evidence to support any *a priori* decisions (Collins & Graham 2002). Repeated assessments have therefore been recommended as a way of best establishing when factors change. Longitudinal designs that incorporate large numbers of assessments through the use of monitoring forms and diary studies allow the collection of a large amount of data on specific questions and help minimise recall bias (Laurenceau et al 2007). They also allow greater flexibility, with a choice of measurement points to better investigate change over time (Collins & Graham 2002). However, repeated assessments increase costs in terms of materials and patient burden which may reduce study uptake and/or increase dropout rates (Laurenceau et al 2007).

The use of diary data, where a small amount of information on each construct of interest is collected several times, may be one way of balancing responder burden and costs with the need to collect data more frequently. Bolger et al (2003) summarise the many advantages of diary studies for investigating how people change over time, the within- and between-person variances in such changes, and the processes underlying these changes. They also emphasise the importance of temporality and how the decision of when to measure particular constructs must be based on theoretical understanding of when the construct of interest is most likely to change. Bolger et al also acknowledge that in many cases there may be little evidence of when constructs may be likely to change, and in these cases advocate *a priori*, well thought-out decisions on when would be the best time to measure specific constructs of interest are required. However, they suggest that more frequent intervals might be better than less frequent intervals, which may lead to greater responder burden.

In summary, the addition of measurements collected during the treatment period has been suggested as a way to improve mediation studies by giving more information on when potential mediators change and whether this is before or after change in the outcome of interest. Using diaries to collect this information, which often contain only single questions per construct, may be a way of collecting information during treatment in a way that leads to minimal increase in responder burden.

8.3 Current Study

The purpose of this analysis was to investigate whether including information on key mediators of interest during treatment as well as pre- and post-treatment can improve future mediation analyses. The objectives were three-fold:

- To investigate at what point potential mediating factors might change (improve), and the type of improvement observed, in order to help inform where the optimal measurement points might be in future mediation studies;
- To investigate whether the inclusion of data measured during treatment provides more information than pre-post analysis, thus helping to answer the question of whether measuring factors of interest during treatment is necessary;
- To compare those who did well (less disability at post-treatment) with those who did not, and investigate to what extent change in potential mediating factors is associated with this change (i.e. what was key/important in leading to improved outcome).

8.4 Study Sample: IMPACT Service

The IMPACT programme is a locally delivered NHS psychological group treatment programme based on Acceptance and Commitment Therapy (ACT), a form of cognitive-behavioural therapy, for people with chronic pain. This treatment is explained in detail in Chapter 1. Patients are either referred by their GP or from secondary care pain or musculoskeletal (MSK) services and therapy is delivered in groups. Diary data on factors relevant to the treatment programme (what the ACT programme is aimed at changing) is collected weekly as part of the patient's therapy, and the scores are discussed as part of team meetings about each patient's individual progress. No data was available on these patients' demographic characteristics from the IMPACT service to preserve confidentiality. Data were available for 80 patients at baseline (pre-treatment) (see Table 8.1). This data is not part of an RCT (so no control group is included), and all patients received the same treatment programme.

In the IMPACT service, patients are seen in groups of a maximum of 12, although some groups may be as small as two patients if dropout occurs over the programme period (J Levell, personal communication). There is no set content for each session, with areas being revisited as needed or as requested by patients. Measures of several of the core processes in ACT were taken, and are described below.

The analyses presented in this chapter utilised data collected between January 2011 and May 2012. Patients in this dataset therefore were part of different treatment groups and started and finished their treatment at different times depending on what group they were in. However, in this analysis the patients were analysed as a single group cohort, as it was not clear which patients belonged to which group. No important changes were made to the programme during the period of time these data were collected so all patients received the same treatment programme. All patients were provided with an information sheet explaining that any information they provided in the questionnaires may be used for research and gave consent for their information to be used in this way. Approval was provided by Staffordshire Research Ethics Committee in 2011 (Ref 10/H1203/57).

Table 8.1 Baseline (pre-treatment) characteristics of study population (*n*=80)

Variable	**Mean (SD) or %**
Completed treatment	96.3% (*n*= 77)
Current Pain at baseline (0-10)	5.9 (1.7)
Average Pain at baseline (0-10)	7.3 (1.59)
Least Pain at baseline (0-10)	4.8 (1.98)
Sickness Impact Profile (SIP) Disability score at baseline (0-1)	0.22 (0.15)

8.4.1 Measures

The measures of interest in this study were disability (outcome) measured at pre-treatment, post-treatment and at three-month follow-up using the Physical Disability subscale of the Sickness Impact Profile (SIP (Bergner et al 1981)), and the five potential mediators which were taken weekly before each treatment session (see Table 8.2). These questions were not taken from a validated questionnaire but instead were devised by the IMPACT team. Some of the potential mediators map directly onto some of the key ACT processes, for example Struggle is a measure of cognitive defusion, Willingness is a measure of acceptance, Values is a measure of valued action and Workability is a measure of committed action.

Table 8.2 Summary details of outcome and process measures

Questionnaire/Sub-scales	**When measured**	**How scored**
Sickness Impact Profile (SIP)	Pre-treatment, post-treatment and three-month follow-up	Scores on each subscale range from 0 to 1 – score closer to 1 indicates greater disability (Vowles et al 2013)
Distress (*Rate how upset and/or distressed you were overall in the past week)*	Pre-treatment, weekly during treatment, post-treatment and three-month follow-up	0-10 Likert scale (0=None/Not at all to 10=Most/Worst
Willingness *(Rate how willing you were to have pain and distress in the past week)*	Pre-treatment, weekly during treatment, post-treatment and three-month follow-up	0-10 Likert scale (0=None/Not at all to 10=Most/Worst)
Struggle *(Rate how much effort you put into making pain or upsetting thoughts, feelings or memories go away this past week)*	Pre-treatment, weekly during treatment, post-treatment and three-month follow-up	0-10 Likert scale (0=None/Not at all to 10=Most/Worst)
Workability *(Rate how effective you were in taking actions that contributed to a better, more vital, quality of living in the past week)*	Pre-treatment, weekly during treatment, post-treatment and three-month follow-up	0-10 Likert scale (0=None/Not at all to 10=Most/Worst)
Valued Action *(Rate how effective you were this past week in making progress in the areas of your life that matter to you)*	Pre-treatment, weekly during treatment, post-treatment and three-month follow-up	0-10 Likert scale (0=None/Not at all to 10=Most/Worst)

8.5 Methodology

All analyses were carried out using SPSS PASW Version 19. Because an increase in distress score originally meant a negative change (more distress) in the dataset, the distress item was reverse scored so that a higher score in distress indicated a positive change (improvement). This is to make the scoring consistent with the other potential mediator variables.

8.5.1 Missing data

Of the 80 patients who initially enrolled in the treatment, only three did not complete treatment. Reasons for this included stress and further investigations for health problems. However, a further six patients did not complete the post-treatment follow-up questionnaires and a total of 32 patients did not complete the three-month follow-up questionnaires (40% of the original sample). Response bias analysis was therefore conducted to see whether those who did not respond at post-treatment and three-month follow-up differed substantially at pre-treatment (baseline) from those who did respond. Descriptive statistics were provided for responders and non-responders at pre-treatment to describe the differences between each group. Comparisons between responders and non-responders were made for baseline disability score and baseline values of depression, pain-related anxiety and each of the potential mediator variables. Depression and anxiety were included because, while these were not key targets for treatment in ACT and therefore not included in the analysis, it was felt that these factors might be important in describing the population.

8.5.2 Descriptive analysis

Descriptive statistics (means and standard deviations (SDs)) were calculated for each potential mediator variable and the outcome measure at pre-treatment, post-treatment and three-month follow-up. This was to see if the measures showed a pattern of change over time. Percentage

change from pre-treatment score was also calculated for each variable to show the proportion of change occurring at post-treatment and three-month follow-up.

8.5.3 Analysing change

This analysis examined change over time to investigate when this occurs and whether the inclusion of assessments during treatment is useful. Leffondré et al (2004) describe 27 different ways of examining change, which they grouped into five categories: *basic descriptives*, including the range and examining means over time; *elementary measures*, which examine change between the first and last score (such as the difference between the first and last score, or the slope in a latent growth model (LGM)); *the examination of non-linearity and inconsistency of change*, which are based on the change between two consecutive scores in the same case (such as SD of the first differences or mean of absolute first differences), with a high value being indicative of non-linearity and inconsistency respectively; *the examination of sensitivity to abrupt short-term fluctuations*, or the difference between two consecutive first differences (i.e. if a large increase in score is followed immediately by a large decrease), which include the maximum absolute second differences (a large value would indicate at least one important abrupt change); and finally *contrasting early and late change*, which compares changes in early and later follow-up (such as the ratio between early and later change, or early and total change). For this analysis, it was decided that descriptive analyses and regression analyses should be produced to look at several different types of change that were identified in the Leffondré paper to address each of the objectives outlined above.

Objective 1 involved classifying patients into groups to investigate different types of change. The categories of *examining non-linearity and inconsistency of changes* and *examining sensitivity to non-monotonicity and to abrupt short-term fluctuations* as described by Leffondré were thought

to be useful to examine, as it was anticipated that the changes across the time points would not necessarily be linear. Leffondré et al (2004) reported 14 different ways of measuring these categories of change, and provided criteria to help decide on the most appropriate measure to use for the present data. The authors also mention *measures contrasting early vs. later change* as important to examine. In the present analysis, cut-offs to describe change were based on clinically relevant cut-offs decided upon by the study team, rather than statistical cut-offs, to examine differences between individuals who changed early or later on in treatment, reported large or small change and reported consistent or inconsistent change. This was because the aim was to investigate clinically relevant change. In addition to this, change between pre- and post-treatment (*elementary change*) was examined to see how many patients reported improvement by post-treatment. How the data were classified is described in Box 8.1 below.

Box 8.1 Types of change in potential mediators examined in the IMPACT dataset

Early vs. Late vs. None vs. Both Early and Late

- 0 = No Improvement (< 1 point positive change at any time point)
- 1 = Early Improvement (≥ 1 point positive change at D1 – PreTreat or D2 – D1)
- 2 = Late Improvement (≥ 1 point positive change at D3 – D2, D4 – D3 or PostTreat – D4)
- 3 = Both Early and Late Improvement (≥ 1 point positive change at D1 – PreTreat, D2 – D1, D3 – D2, D4 – D3 or PostTreat – D4)

Large vs. Small (change of 2 points at any point over the 4-week period)

- 0 = Small Improvement (< 2 points positive change at any time point, includes no change and worsening)
- 1 = Large Improvement (≥ 2 points positive change at any time point)

Consistent vs. Not Consistent (for at least 2 time points)

- 0 = Not Consistent Improvement (< 1 point positive change or ≥ 1 point positive change for less than 2 consecutive time points – includes worsening)
- 1 = Consistent Improvement (≥ 1 point positive change for ≥ 2 consecutive time points – no worsening)

Improved vs. Worsening vs. No Change (pre-post)

- 0 = No Change (< 1 point at any time point)
- 1 = Worsening (≥ 1 point decrease between pre- and post-treatment (negative change))
- 2 = Improvement (≥ 1 point increase between pre- and post-treatment (positive change))

D1-D4= Diary assessment points weeks 1 to 4 (during treatment)

Objective 2 involved comparing mean changes between pre-and post-treatment assessment with changes observed during treatment. Means and SDs were calculated for each measure at each available time point to see whether there was change over time for the cohort as a whole.

Individual scores were also plotted for each variable to get a sense of the amount of between-patient variability in scores pre- and post-treatment and also during treatment. It was hoped that LGMs could be built here to make use of all the available time points, but the available sample size in the present analysis was too small (see Section 8.7 for further discussion).

Objective 3 examined whether the potential mediators of interest were associated with improvement in outcome, as measured by residualised change in the SIP physical function subscale. Although this cannot establish whether the variable was a mediator, it will clarify its potential to mediate outcome by testing whether change in that variable was associated with change in outcome. Univariable linear regression analyses were used to compare how the different measures of change used in Objective 1 described the association between each of the potential mediators and change in the outcome (disability). As linear regression assumes normal distribution in the outcome variable, this was tested prior to the analysis. The % R^2 change, *F* ratio and beta (*b*) are the statistics of interest in this analysis. These statistics give an indication of the amount of variance explained by the model (%R^2), whether the model provides a good fit to the data (*F* ratio) and the strength of the relationship between the predictor and outcome (*b*) (Field 2009).

8.6 Results

8.6.1 Missing data

Table 8.3 shows the means and SDs for baseline characteristics of those who responded and did not respond at post-treatment and three-month follow-up. There were significant differences between the groups, with non-responders being more physically disabled, depressed, and anxious about their pain compared to responders at both post-treatment and three-month follow-up. This

demonstrated the likelihood of a degree of bias in the results, meaning that the findings may not be applicable to all patients with chronic pain, particularly those who are less likely to have a favourable outcome.

Within the dataset there was also a substantial amount of missing data for a particular item in the diary study (Willingness), with between 25% and 56% missing data at each assessment point. On discussion with members of the IMPACT study team it was found that many patients struggled with the Willingness question, even by the end of treatment, and preferred not to give an answer to this question (J Levell, personal communication). The decision was therefore made to not include this particular item in the analysis, due to the small number of cases who had responded and the likely bias in these responses.

Table 8.3 Baseline characteristics (Means and SDs) of responders versus non-responders at post-treatment and three-month follow-up

Measure	**Baseline assessment (Mean (SD)) (*n*=80)**	**Post treatment Baseline values (Mean (SD))**		**Three month follow-up Baseline values (Mean (SD))**	
		Resp (*n*=69)	**Non-resp (*n*=10)**	**Resp (*n*=48)**	**Non-resp (*n*=30)**
Physical Disability (SIP Subscale) Range 0-1	0.2 (0.2)	0.22 (0.2)	0.3 (0.2)	0.2 (0.2)	0.3 (0.2)
Depression (British Columbia Major Depression Inventory (BCMDI)) Range 0-80	33.3 (12.4)	33.4 (11.9)	32.8 (15.5)	32.2 (11.0)	34.8 (14.2)
Pain-related Anxiety (Pain Anxiety Symptoms Scale (PASS)) Range 0-100	48.8 (19.0)	49.1 (18.3)	46.7 (24.1)	49.7 (18.9)	47.5 (19.3)
Distress (0-10)	3.2 (2.5)	6.9 (2.5)	6.3 (2.5)	6.8 (2.6)	6.9 (2.4)
Struggle (0-10)	7.1 (2.6)	7.2 (2.5)	6.6 (3.0)	6.5 (2.8)	7.9 (2.0)
Workability (0-10)	5.7 (2.4)	5.6 (2.4)	6.0 (2.9)	5.8 (2.2)	5.5 (2.7)
Valued Action (0-10)	4.8 (2.7)	4.8 (2.7)	4.6 (3.1)	4.8 (2.8)	5.0 (2.7)

For the remaining data in the study, cases were classified as 'missing' if no data was available for a particular person at any follow-up time point. Cases were also classified as missing if less than one score was available for Early vs. Late Improvement and Small vs. Large Improvement (i.e. no score available for the first or last two measurements taken during treatment), or if less than two scores were available for Non-Consistent vs. Consistent Improvement. This was because for the first two types of change, at least one score during treatment is needed to see where change occurred and how large this was, and for the last type of change two scores were needed to establish whether the change was consistent.

8.6.2 Descriptive analysis

Table 8.4 shows the mean scores and SDs at each assessment period for the potential mediators and outcome measures. The proportion of change in the variable (follow-up score minus pre-treatment score) is also expressed as percentage change. The percentages show that for Struggle, there was very little change at any time point; for Workability the largest proportions of change occurred at weeks 2 and 4 of treatment; and for Distress and Valued Action the most change occurred in the final weeks of treatment. Disability (measured only at pre- and post-treatment) improved by 27%.

Table 8.4 Mean score (SD) and proportion change (%) at each assessment period for the outcome (SIP) and each potential mediator variable

		Pre-treatment (T2) Mean and SD	Week 1 (D1) Mean (SD)	Proportion of change from pre-treatment (%)	Week 2 (D2) Mean (SD)	Proportion of change from pre-treatment (%)	Week 3 (D3) Mean (SD)	Proportion of change from pre-treatment (%)	Week 4 (D4) Mean (SD)	Proportion of change from pre-treatment (%)	Post-treatment (T3) Mean (SD)	Proportion of change from pre-treatment (%)
SIP Disability subscale (0-1)	Higher score = less disability (reverse scored)	0.22 (0.15)	***	***	***	***	***	***	***	***	0.16 (0.14)	27.2%
Distress (single item diary question – 0-10 scale)	Higher score = better functioning (reverse scored)	3.2 (2.47)	3.4 (2.70)	6.3%	3.4 (2.58)	6.3%	4.0 (2.50)	25%	5.0 (2.93)	56.3%	5.4 (2.70)	68.8%
Struggle (single item diary question – 0-10 scale)	Higher score = more effort (positive)	7.1 (2.56)	7.0 (2.58)	1.4%	6.8 (2.20)	4.2%	6.4 (2.53)	9.9%	6.4 (2.80)	9.9%	6.6 (2.59)	7.0%
Workability (single item diary question – 0-10 scale)	Higher score = more effective	5.7 (2.44)	5.0 (2.53)	12.3%	5.3 (2.40)	7.0%	5.8 (2.26)	1.8%	6.5 (2.37)	14.04%	7.0 (2.30)	22.8%
Valued Action (single item diary question – 0-10 scale)	Higher score = more progress	4.8 (2.72)	4.7 (2.49)	2.1%	5.4 (2.17)	12.5%	6.3 (2.06)	31.3%	6.9 (2.26)	43.8%	7.0 (2.35)	45.8%

****=not measured*

In Objective 1 the time at which change occurred in individuals who took part in the IMPACT ACT therapy was investigated by classifying individuals into groups to look at different types of change (see Box 8.1). For all of the potential mediator variables, change occurred later rather than earlier for the majority of patients (between 29% and 43% for the four variables - see Table 8.5) although change in Distress, Workability and Valued Action often occurred both early and later on (32.5% and 32.5% respectively for Distress; 38.8% and 28.8% respectively for Workability; 38.8% and 33.8% respectively for Valued Action). A large change (improvement of two points or more) occurred for the majority of patients at some point during the treatment period (between 48% and 71% for the four variables), except for Struggle where there were similar numbers of patients experiencing large and small change (48% and 50% respectively). Change tended to be inconsistent for the majority of patients across each of the variables (between 58% and 89%). Finally, the percentages of those classified as Improved, Not improved or No change pre-post treatment suggested that there was improvement by post treatment (between 23% and 56% for the four variables), although the large amount of missing data for the post-treatment assessment (only *n*=50 remaining) could preclude any definitive conclusions for this final type of change. These changes are represented graphically in Figures 8.1a to 8.1d.

Figure 8.1a Percentages of participants classified as Early or Late improvement across the potential mediator variables

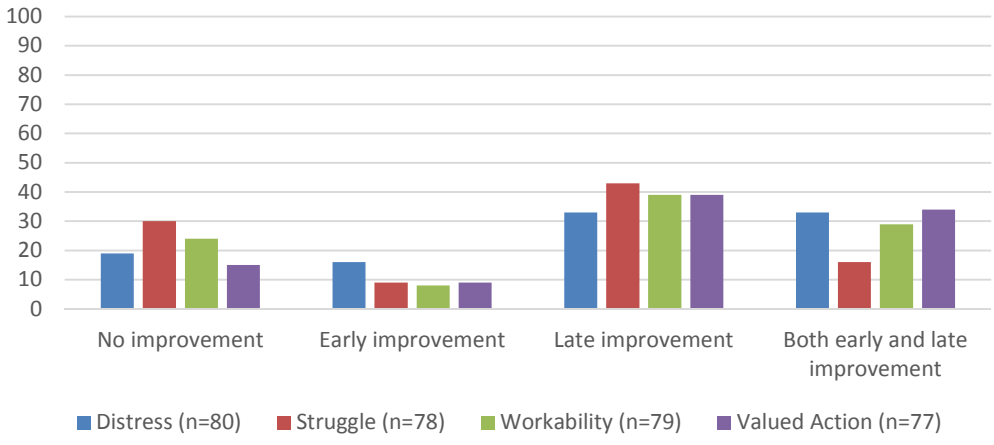


Figure 8.1b Percentages of participants classified as Small or Large improvement across the potential mediator variables

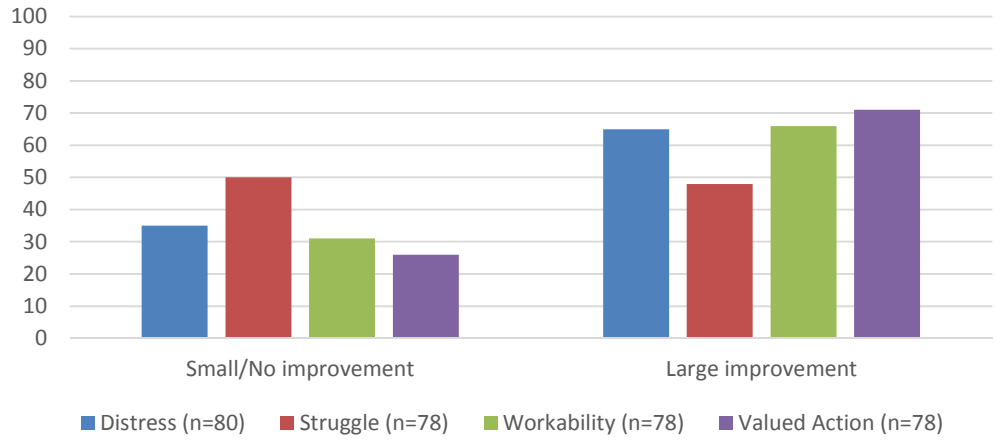


Figure 8.1c Percentages of participants classified as Consistent or Inconsistent change across the potential mediator variables

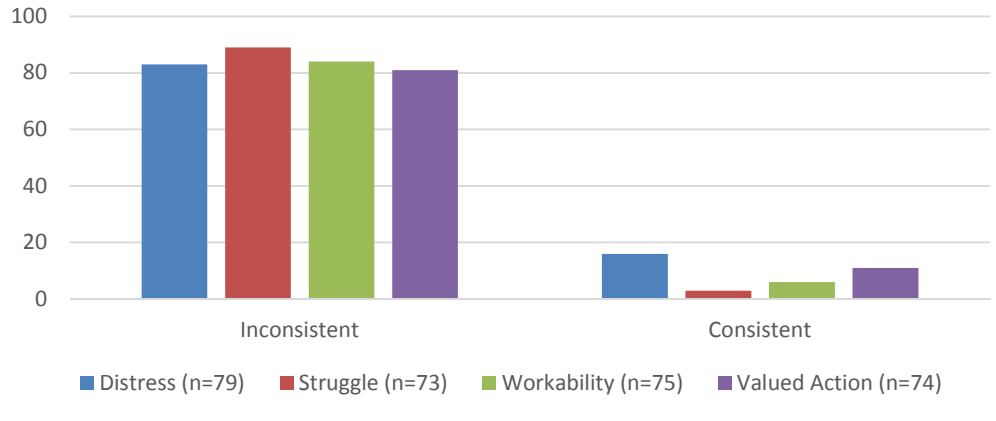
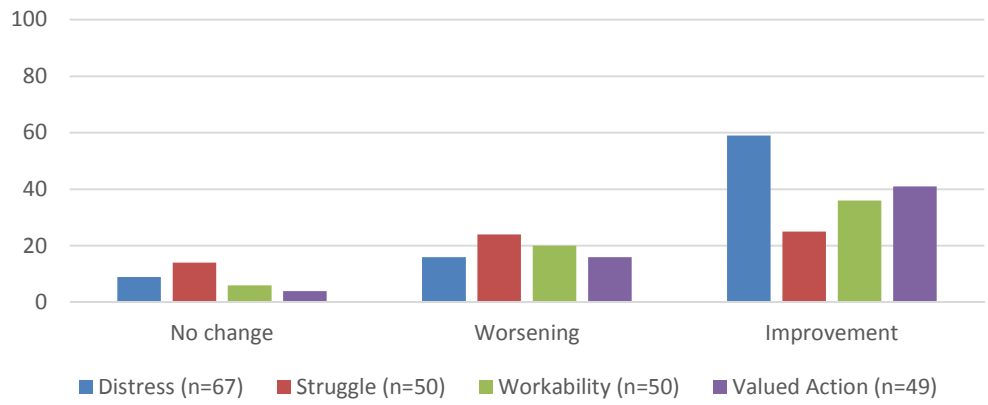


Figure 8.1d Percentages of participants classified as Improved or Not improved across the potential mediator variables



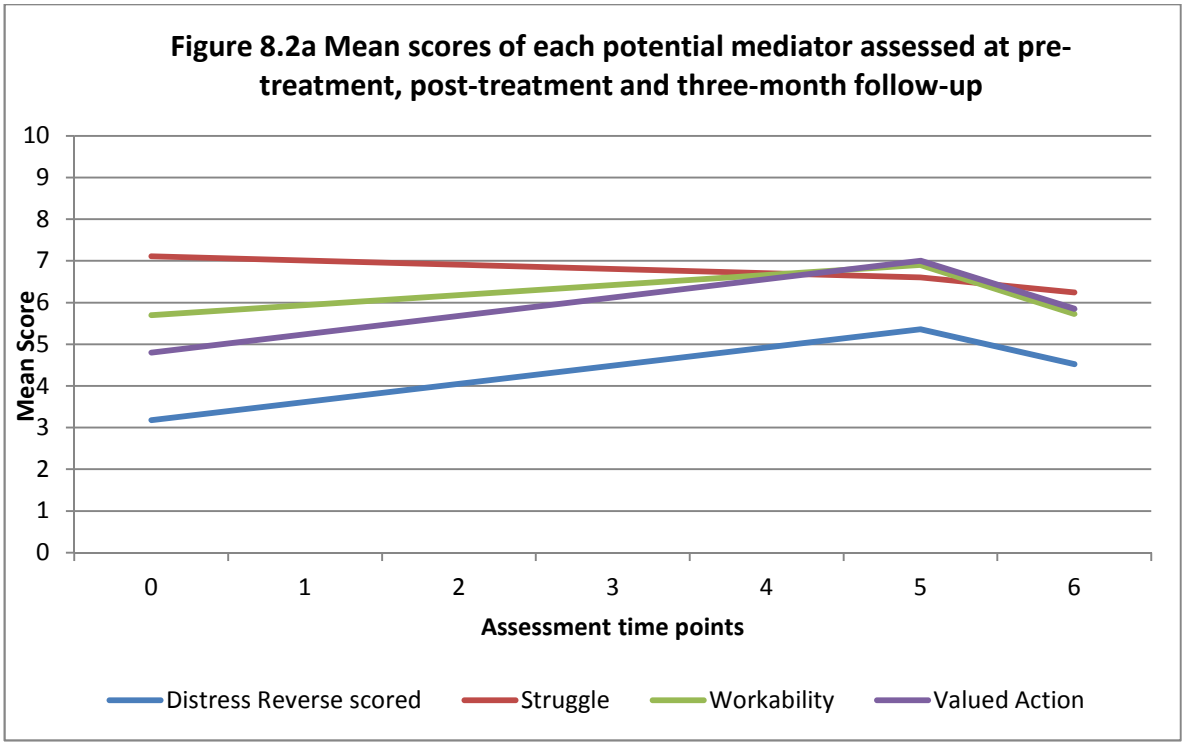
Repeated measures ANOVAs were also conducted to formally test for statistically significant changes over the within-treatment period for each of the potential mediator measures. It was found that Distress ($F_{(3, 183)}$=10.57, p<0.05), Workability ($F_{(3, 183)}$=7.67, p<0.05) and Valued Action ($F_{(3, 183)}$=16.27, p<0.05) significantly improved during the treatment period, but Struggle ($F_{(3, 180)}$=1.67, p=0.17) did not.

Table 8.5 Percentages of classification of change for individual patients in the IMPACT treatment cohort

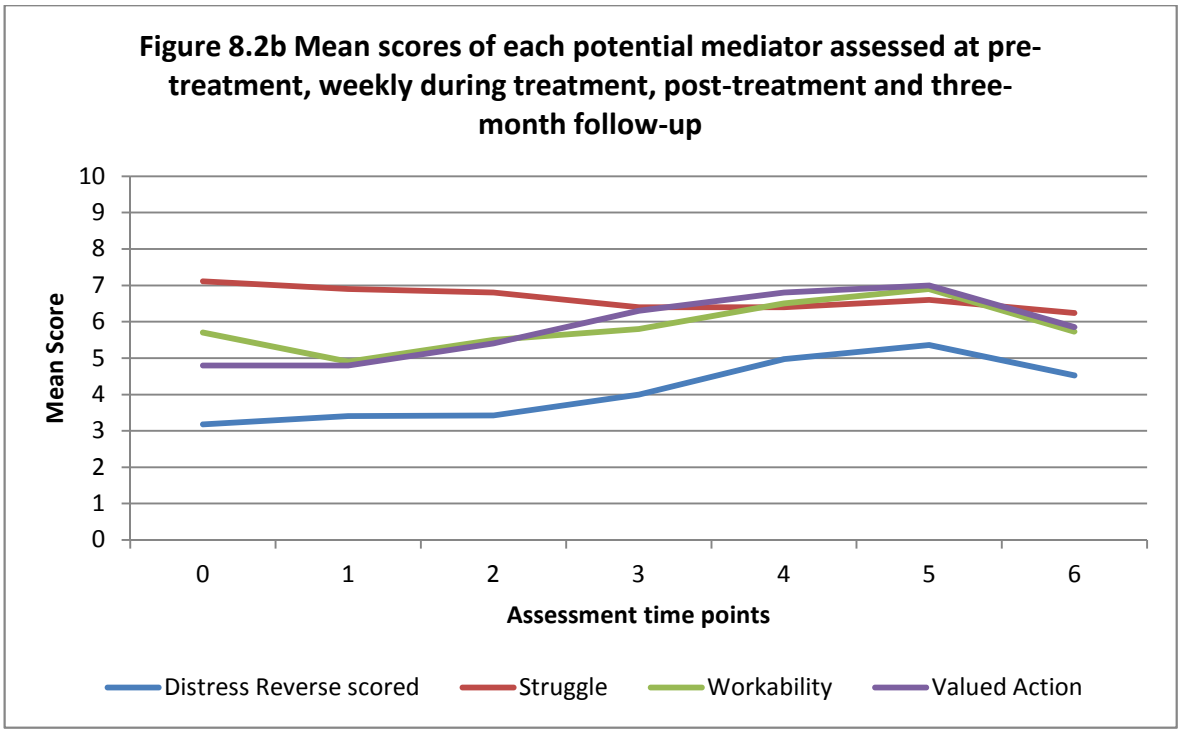
Type of Change		**Distress *n* (%)**	**Struggle *n* (%)**	**Workability *n* (%)**	**Valued Action *n* (%)**
Early vs. Late	No Improvement	15 (18.8%)	24 (30.0%)	19 (23.8%)	12 (15.0%)
	Early Improvement	13 (16.3%)	7 (8.8%)	6 (7.5%)	7 (8.8%)
	Late Improvement	**26 (32.5%)**	**34 (42.5%)**	**31 (38.8%)**	**31 (38.8%)**
	Both Early and Late Improvement	**26 (32.5%)**	13 (16.3%)	23 (28.8%)	**27 (33.8%)**
	% missing	0	2.5%	1.3%	3.8%
Small vs. Large	Small or No Improvement	28 (35.0%)	**40 (50%)**	25 (31.3%)	21 (26.3%)
	Large Improvement	**52 (65.0%)**	**38 (47.5%)**	**53 (66.3%)**	**57 (71.3%)**
	% missing	0	2.5%	2.5%	2.5%
Non-Consistent vs. Consistent	Non-Consistent Improvement	**66 (82.5%)**	**71 (88.8%)**	**70 (83.5%)**	**65 (81.3%)**
	Consistent Improvement	13 (16.3%)	2 (2.5%)	5 (6.3%)	9 (11.3%)
	% missing	1.3%	8.8%	6.3%	7.5%
Improved vs. Not Improved	No Change	7 (8.8%)	11 (13.8%)	5 (6.3%)	3 (3.8%)
	Worsening	13 (16.3%)	**19 (23.8%)**	16 (20.0%)	13 (16.3%)
	Improvement	**47 (58.8%)**	**20 (25.0%)**	**29 (36.3%)**	**33 (41.3%)**
	% missing	16.3%	37.5%	37.5%	38.8%

These analyses demonstrated that while patients tended to report at least one large change over the treatment period and improved on each of the variables by the end of treatment, this change tended to happen later on in the treatment period but was often inconsistent over time. The inconsistency could be indicative of error in the measures, and/or that the patients scored the measures arbitrarily. However, the potential mediators did show similar patterns across the different types of change, suggesting that the treatment delivered in the IMPACT programme was affecting each of the variables in a similar way.

In Objective 2 the question of whether including data collected during treatment provided more information than using only pre- and post-treatment data was investigated. Descriptive analysis was firstly conducted, plotting the means at each time point. Figure 8.2a shows mean scores at pre- and post-treatment only and Figure 8.2b includes the data obtained during treatment. The graphs show how change occurs, at group level, during the treatment period. Overall, Distress had its biggest deterioration between weeks 2 and 3 of treatment, Valued action and Workability improved gradually throughout treatment, and Struggle deteriorated slightly during treatment and by post-treatment. The potential mediator measures indicated that change is not a straightforward process, and that some measures may take longer to shift than others.



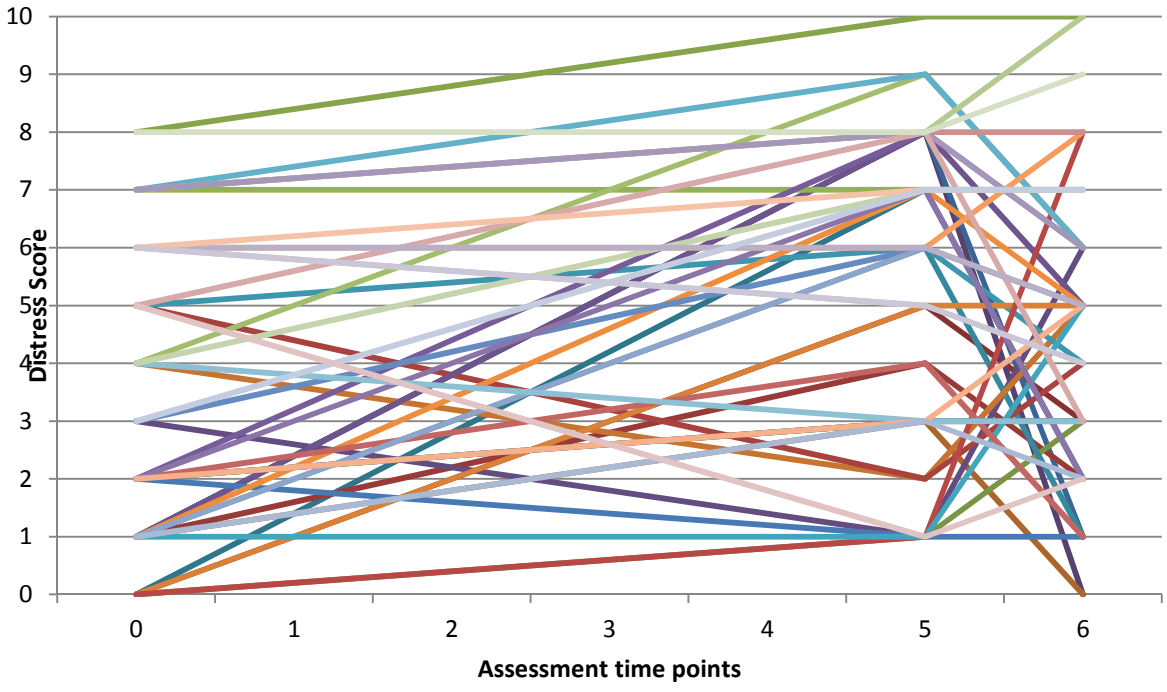
Key: 0=Pre-treatment; 1-4=Diary data; 5=Post-treatment; 6=three-month follow-up



Key: 0=Pre-treatment; 1-4=Diary data; 5=Post-treatment; 6=three-month follow-up

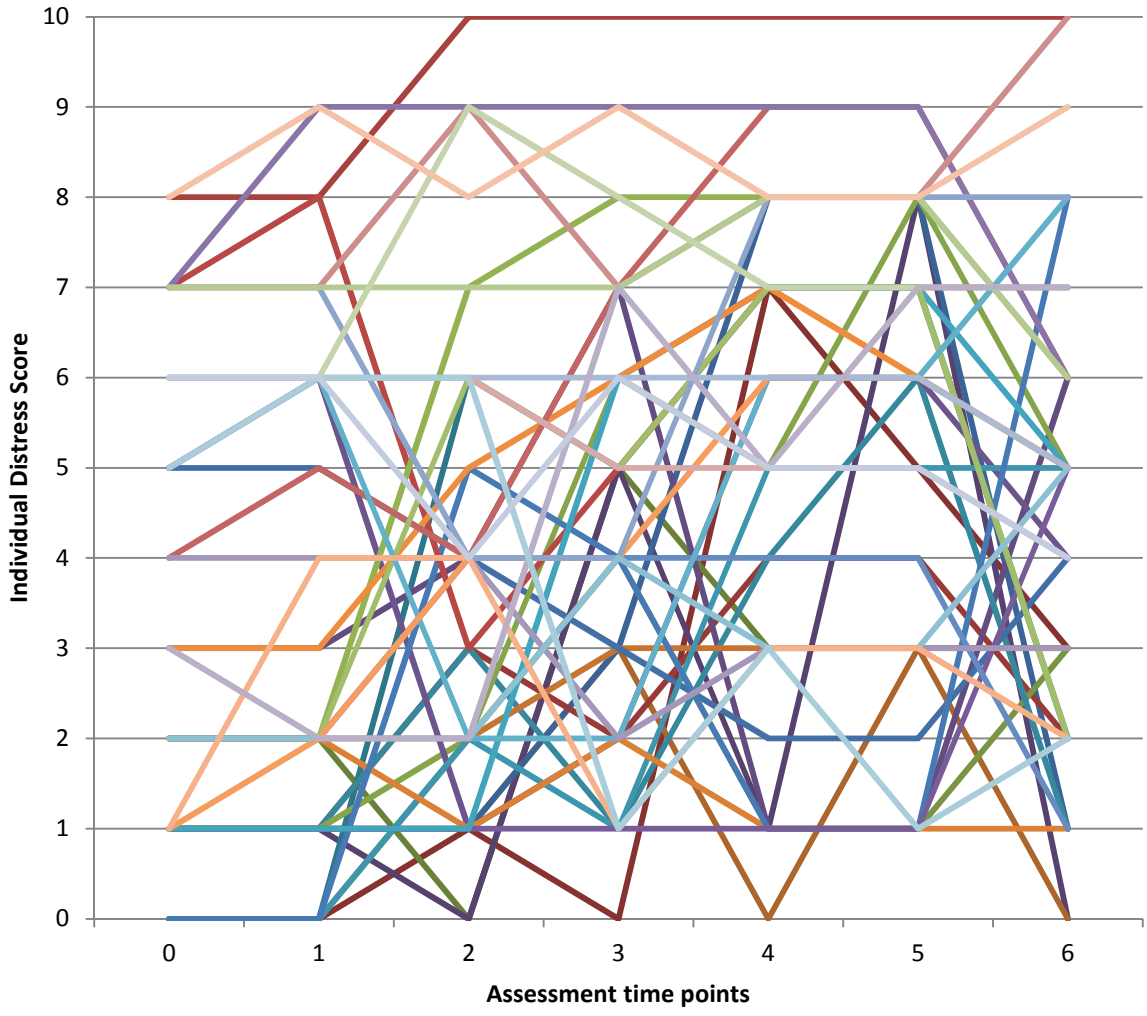
An important point to note is that these changes only explored mean change, and did not allow for the investigation of change in individual patients. Data for individual patients was therefore graphically described for Distress as an example of the individual variation that occurred during treatment. Figures 8.3a and 8.3b below show how pre-post measures masked considerable individual variability among patients, by assuming linear change between pre- and post-treatment which in reality is non-linear. Only complete case data (where data was available at each included assessment point for a particular patient) was included here, hence the smaller numbers represented in Figures 8.3a and 8.3b.

Figure 8.3a Individual patient Distress scores at Pre-treatment, Post-treatment and three-month follow-up (*n*=48)



Key: 0=Pre-treatment; 1-4=Diary data; 5=Post-treatment; 6=three-month follow-up

Figure 8.3b Individual patient Distress scores at Pre-treatment, weekly during treatment, post-treatment and three-month follow-up



Key: 0=Pre-treatment; 1-4=Diary data; 5=Post-treatment; 6=three-month follow-up

Finally, in Objective 3 each of the different types of within-treatment change was compared using linear regression to explore how the association between each potential mediator and the outcome was best described for each variable. Table 8.6 shows the results of the linear regression analyses to compare the different types of change. For this analysis, residualised change in SIP disability score between pre- and post-treatment was used as the outcome. This was because post-treatment SIP disability was not normally distributed, and the use of residualised change improved this (see Appendix 8.1). In the analysis of Early vs Late improvement, all analyses were

compared with the ‘no improvement’ category. The results showed that small proportions of variance in residualised change in disability were explained by early, late or early and late change in Distress, Workability and Valued Action. However, early and late improvement in Workability were both significantly associated with residualised change in SIP Disability score at post-treatment, suggesting that improvement is partly explained by early or late change in this potential mediator. Worsening in valued action compared to no change was also found to be a significant predictor of residualised change in disability according to the 95% CI. However, none of the other types of change in potential mediators examined were shown to explain a significant proportion of the variance or were found to be significantly associated with residualised change SIP Disability score at post-treatment.

Table 8.6 Univariable linear regression analyses comparing the predictive value of each type of within-treatment change on predicting SIP Disability score at post-treatment

Outcome: Indicators of residualised change in SIP Disability Score (Pre-treatment to Post-treatment)					
Type of Change	**Predictor in model**	***% R^2***	***F***	***b* (95% CI) (Unstandardised)**	
Early vs Late Improvement (reference: No improvement)	Distress	8.8%	1.85 (ns)	Early improvement	-0.06 (-0.98 to 0.85)
				Late improvement	-0.64 (-1.40 to 0.12)
				Both early and late improvement	-0.68 (-1.44 to 0.08)
	Struggle	0.7%	0.15 (ns)	Early improvement	0.24 (-0.84 to 1.33)
				Late improvement	-0.04 (-0.62 to 0.53)
				Both early and late improvement	0.12 (-0.61 to 0.85)
	Workability	8.8%	2.04 (ns)	Early improvement	-0.59 (-1.61 to 0.43)
				Late improvement	-0.71 (-1.36 to -0.06)
				Both early and late improvement	**-0.79 (-1.48 to -0.11)**
	Valued Action	1.3%	0.27 (ns)	Early improvement	0.05 (-1.05 to 1.15)
				Late improvement	-0.20 (-1.02 to 0.61)
				Both early and late improvement	-0.28 (-1.10 to 0.55)
Small vs Large Improvement (reference: Small improvement)	Distress	5.3%	3.65 (ns)	Large improvement	-0.51 (-1.05 to 0.02)
	Struggle	0%	0.02 (ns)	Large improvement	0.04 (-0.44 to 0.52)
	Workability	3.1%	2.01 (ns)	Large improvement	-0.38 (-0.92 to 0.16)
	Valued Action	0%	0.01 (ns)	Large improvement	-0.03 (-0.60 to 0.55
Consistent vs Inconsistent Improvement (reference: Not consistent)	Distress	0.5%	0.32 (ns)	Consistent improvement	-0.18 (-0.81 to 0.46)
	Struggle	0.3%	0.16 (ns)	Consistent improvement	0.28 (-1.14 to 1.70)
	Workability	0%	0.02 (ns)	Consistent improvement	-0.07 (-1.09 0.95)
	Valued Action	0.1%	0.04 (ns)	Consistent improvement	0.08 (-0.67 to 0.82)

Outcome: Residualised change in SIP Disability Score (Pre-treatment to Post-treatment)					
Type of Change	**Predictor in model**	***% R^2***	***F***	***b* (95% CI) (Unstandardised)**	
Improvement vs Worsening (Reference: No change)	Distress	3.3%	1.03 (ns)	Worsening	0.55 (-0.48 to 1.59)
				Improvement	0.14 (-0.79 to 1.06)
	Struggle	6.7%	1.58 (ns)	Worsening	-0.29 (-1.04 to 0.47)
				Improvement	-0.63 (-1.39 to 0.12)
	Workability	5.6%	1.29 (ns)	Worsening	0.64 (-0.32 to 1.61)
				Improvement	0.24 (-0.67 to 1.15)
	Valued Action	1.4%	3.09 (ns)	**Worsening**	**1.21 (0.04 to 2.38)**
				Improvement	0.59 (-0.50 to 1.68)

8.7 Discussion

8.7.1 Summary

This descriptive analysis investigated whether including information collected during treatment could be used to improve future mediation analysis studies, as this recommendation has often been made in the mediation literature. Clinical data was provided from a tertiary treatment service implementing group ACT therapy for patients with chronic pain to improve physical functioning. As well as full baseline (pre-treatment) and post-treatment data on physical functioning, data were also collected on factors that the therapy targeted for change at pre-treatment, post-treatment and also at weekly intervals as part of treatment. These factors were therefore labelled as potential mediators in the analysis, although formal mediation analysis was not conducted due to the small number of cases available and the lack of a control group.

Change over time in the potential mediators and outcome variable was firstly described to show whether change occurred over time and the magnitude of that change (expressed as a percentage). This showed that the factors clearly differed on when they changed during treatment, with some changing early on, some changing later and others not changing at all. While this was based on mean scores, this provides a helpful comparison with the ACT theoretical model and clinical opinion on when these factors are expected to change. The data was then classified into different types of change, based on consensus within the research team on what type of change would be clinically useful to examine. The frequencies of these different types of change for each of the potential mediator variables showed that most patients experienced at least one large change (of more than two points) on all factors and that most experienced improvement, but that improvement was inconsistent. This could be due to the measures used in the analysis (see further discussion in Section 8.7.4 below).

Next, comparisons were made between pre-post assessment only and the inclusion of assessments made during treatment. The means of each variable were first plotted which showed some differences between measurement at pre-post only and the inclusion of mid-treatment scores, but when individual patient scores were plotted this showed a much larger variability during treatment. This highlighted the need to also include analysis that could adequately incorporate this individual-level change as well as group-level change.

Finally, the association of the different types of change with residualised change in physical function were investigated to explore whether a particular type of change may be important in leading to a change in outcome. This analysis showed that very little variance of change in physical function was explained by any of the types of change described. This suggests that none of these measures of change adequately reflect what occurs during treatment, or that change in the mediating factors during treatment is not strong enough to change disability outcome. It could also be a reflection of the small sample size available for analysis, which limits the ability to detect relevant difference and also precluded a stronger statistical analysis method such as LGM.

8.7.2 Comparison with conceptual model

ACT involves building acceptance and psychological flexibility using metaphors and stories to help patients apply the therapeutic concepts. The six processes outlined in Chapter 1 are described as 'overlapping' and in practice are described as flexible (Hayes et al 2013). In the IMPACT service, concepts were revisited as and when needed and therapists expect an increase in Struggle initially and large changes at points where patients start to understand and apply the concepts, which are likely to vary from patient to patient (J Levell, personal communication). The results presented here are consistent with the expectation that large changes occur for most patients and change for most of the potential mediators occurred towards the end of the treatment period, which

again appears to be consistent with clinical expectations. However, the finding that Struggle in particular changed very little over time could suggest that the therapy is not targeting this variable strongly, or perhaps that it is not being captured well by the measure (see Section 8.7.4).

8.7.3 Comparison with previous findings

Vowles et al (2013) also carried out analysis on data collected from the IMPACT service with the aim of identifying treatment mediators of disability outcome. This study did not include any of the mid-treatment measures, and used full scale measures of variables related to ACT at pre- and post-treatment. Multi-level structural equation modelling (SEM) was used to take account of the fact that patients attended therapy in groups. Statistically significant mediating effects were found for almost all of their included potential mediators (pain acceptance, psychological flexibility, self-compassion and values) on disability as a latent variable (including observed measures of self-reported and observed physical function measures) but when physical disability (SIP Physical Disability subscale) only was investigated, only pain acceptance was found to have a statistically significant mediating effect with very small (-0.001) *b* path coefficients contributing to this. This highlights potential issues with this measure being used as an outcome.

8.7.4 Limitations of the present study

These analyses focused on exploratory, descriptive analysis because the small sample size available precluded more complex analysis. For example, it was originally planned that LGM could be performed to better compare change that occurred pre- and post-treatment with change that incorporated the mid-treatment assessments, but problems with model fit and the statistics produced indicated that the sample size was not large enough for robust results to be produced. The measurement of the outcome of interest only at pre- and post-treatment and not during treatment also precluded the use of potential alternative analyses, such as Generalised Estimated

Equations (GEE), which also allows for multiple measurement and is specifically designed for use of repeated measures of non-normal data (Ballinger 2004). However, while the analyses presented in this chapter are less technical than in the previous chapters, important information can still be derived on when factors change that may be useful to the IMPACT team, and highlight the huge individual variability in patients and the need to account for this in future studies.

The data collected in this study were based on self-report, which could be problematic in this study for a number of reasons. Firstly, as the data were collected as part of treatment and scores were discussed during treatment with the psychologist and other group members, this may have had an impact on what score patients' chose to give (social desirability bias (Streiner & Norman 2008)), and also patients previous scores may influence subsequent ones, as weekly time intervals may be short enough to allow patients to recall their previous score.

It is also likely that the scores given on each of the questions represent more than a response to the question being asked – wide fluctuations in score across the treatment period for the same patient may reflect disengagement or frustration, for example, rather than actual feelings of 'valued action'. Related to this is the measures themselves – each variable was measured via a single item on a Likert scale, and while it is understandable that short measures were needed to reduce responder burden, the complexity of these concepts is unlikely to be adequately measured by a single question (for example, the question "Rate how effective you were in taking actions that contributed to a better, more vital, quality of living in the past week" is complex, and patients may have succeeded or struggled on different aspects or actions over the course of a week). This highlights a tension between being able to include and assess variables of interest adequately and repeatedly in a study while not overburdening patients or inadvertently introducing bias.

A final issue regarding the items used in the analysis is that they were not tested adequately for validity or reliability in the patient population of interest. These questions were used to collect data for clinical use primarily, but also for use within research studies. In line with the analyses presented in Chapters 5-7 in this thesis, a future recommendation is that strong measures are used to collect data that are comprehensible to patients and can be shown to have sound psychometric properties.

Another issue related to measurement is that of correlations between repeated measures and also errors in those measurements, especially when measures are taken at short time intervals (Gibbons et al 2010). The descriptive data and linear regression analyses presented here cannot take account of this, but the other complex methods mentioned above (LGM, GEE) can account for this correlation (Gibbons et al 2010) and should therefore be considered for use in future studies.

8.7.5 Clinical implications

Clinically, these results showed that while most of the potential mediator variables improved during treatment, Struggle did not appear to change. This variable may be stable because it is not be targeted optimally by the ACT programme or because this particular factor is not amenable to change. It could also be that patients are not comfortable answering this question or that this measure is not particularly useful to include in treatment. The fact that change tended to occur later on for most of the variables is consistent with the IMPACT team's expectations and with the ACT theoretical model, providing preliminary support for this model.

In order to improve the quality and robustness of future analysis, more data and a larger sample size is required for more sophisticated analyses to be employed. Full measures of each of the

potential mediator variables, measured at the time when these variables are most likely to change, may be more beneficial than repeated shorter measures to capture complex concepts such as those used in ACT, or at the least developmental work should be done to ensure that answers reflect the concept being measured and not other aspects of treatment.

8.8 Conclusion

Insight into when variables of interest change was gained from descriptive analyses of data collected during treatment. This may help refine future studies of ACT in terms of when the optimal measurement points for specific variables would be. The individual variation seen in the measures of interest over time also underlines the importance of choosing analysis methods that can account for these individual differences. However, these analyses were preliminary and several issues with the measurement of the potential mediators and outcome used in this study means that further work is needed to confirm the usefulness of including measures during the treatment period.

Chapter 9. Overall Discussion

9.1 Summary of Thesis

This thesis has focused on the challenges of designing studies to analyse mediators of psychological treatments for patients with musculoskeletal (MSK) pain. Chapter 1 described the rationale for focusing on this population, describing MSK pain (and low back pain (LBP) specifically) as being a considerable burden in terms of its effect on quality of life and the financial impact on the NHS and workplaces. It was clear that identifying which factors (mediators) change during treatment and lead to improvement in functional outcomes might be an important step towards improving treatment effects. Psychological factors, the focus of this thesis, were found to be predictive of functional outcomes in observational studies of patients with MSK pain, suggesting they may also be useful factors to target during treatment. However, psychological treatments have only been found to have small or no effect on outcomes, calling into question whether psychological factors can be modified in a way that impacts on treatment outcome and if so, how this can be done to maximise their effect. Mediation analysis was therefore explored as a way of testing this. The overall aim and specific objectives of the thesis are described in Chapter 2.

Chapter 3 outlined the current literature on mediation analysis, both within and outside of the MSK pain literature, in order to summarise current best practice for the design and analysis of mediation studies. Both design and analysis issues with current mediation practice were discussed. This chapter highlighted the complexities of mediation analysis, and that even though the question of 'what makes a treatment work' is on the surface a simple one, actually establishing what was indeed responsible for the success (or failure) of a treatment requires testing of causal associations, which is optimally done via a programme of research which builds up evidence for factors to be tested as part of a randomised controlled trial (RCT). However, a

number of recommendations were also derived from the literature about the best research design to test treatment mediation and the best methods for conducting the analysis. The aim of the research conducted in this thesis, then, was to evaluate methods of treatment mediation analysis in MSK pain populations, starting with a systematic review of previous treatment mediation analysis of psychological treatments for MSK pain and following on with secondary analysis conducted in different study designs and settings in order to test some of the recommendations made in the literature review.

The systematic review presented in Chapter 4 identified 10 papers, three of which focused on LBP populations, which had carried out treatment mediation analysis of a psychologically-focused treatment aimed at reducing disability in patients with MSK pain. While several factors were identified as mediators, including pain catastrophising and pain self-efficacy, the design and methodology of each of the studies contained many of the flaws identified in Chapter 3, which led to a lack of confidence in the results overall. Such issues included a lack of clear theoretical basis for the intervention, not using strong measures of the psychological factors being tested as mediators or not reporting the psychometric properties of the measures used, and using substandard methods of mediation analysis. The findings were consistent with systematic reviews of treatment mediation in RCTs from other fields, including nutrition and oncology, which also reported problems with study design and analysis. This precludes strong conclusions as to what factors actually mediated outcome. The statistical mediation analysis presented in this thesis therefore (as far as possible) followed the recommendations made by the literature and systematic reviews in order to produce robust mediation analysis.

In Chapter 5, the use of observational data to identify potential treatment mediators by generating information on which psychological factors were most likely to be modifiable and

change over time, and therefore most likely to potentially mediate treatment outcomes, was considered. Although not mediation analysis as such, the approach is in accordance with the recommendation of using existing evidence and building on non-intervention study data to identify factors to be tested as treatment targets. The analysis presented in this chapter was secondary analysis of the BeBACK study, which had previously identified four factors (self-efficacy and three different illness perceptions) as being independently predictive of disability outcome. However, the analysis presented in Chapter 5 investigated specifically whether the variables changed over time and were related to change in outcome, and the analysis revealed that not all of those factors found to be predictive in the original study were also potential mediators of disability outcome, demonstrating that just because a variable is a strong predictor of outcome, does not mean it is necessarily modifiable and related to change in that outcome. Four factors in particular (self-efficacy, anxiety, depression and pain catastrophising) were found to change over time, and change in each factor was strongly related to change in outcome. Three of these factors were then tested as mediators of a specific treatment effect in Chapter 6, providing an example of how observational study data can be used to identify potential treatment targets.

In Chapter 6 secondary analysis of the STarT Back trial data was presented in order to test the trial authors' hypothesis that the treatment, which involved patients with LBP who were at high risk of developing long-term disability receiving psychologically informed physiotherapy to target obstacles to recovery (if allocated to the treatment arm), successfully reduced disability, an important aspect of which was to reduce patients' psychological distress. This analysis comprised mediation analysis using the recommended method (product of coefficients approach with bias-corrected bootstrapped 95% CIs to ascertain the magnitude of the mediated effect). The study found that change in psychological distress was a strong mediator of both change in disability and general physical health (the functional outcomes of interest available in this dataset). There were

methodological limitations in the design of this study, such as the control group receiving treatment that might also have influenced distress, 'psychological distress' possibly being too broad a target, the potential for the results being affected by attrition bias, and assessments not being taken at enough time points to establish a causal relationship between the mediator and outcome. The latter suggests that interventions that are targeted to affect a specific factor, with at least three assessment points included in the analysis, may be a better design to test for mediating effects.

Further secondary analysis was therefore conducted with another RCT of a psychologically focused intervention for patients with LBP, which addressed some of the limitations present in the analysis in Chapter 6 and met the criteria of targeting a specific factor (fear-avoidance beliefs) and assessing variables across more than three points in time (five assessment points were included). The secondary analysis of the Back In Action study presented in Chapter 7 also offered the opportunity to test an alternative, more novel form of mediation analysis seen as an important future direction in Chapter 3 (latent growth modelling (LGM)). This analysis found that change in fear-avoidance beliefs was found to be a strong mediator of change in disability in an intervention study that was designed to assess and address fear-avoidance beliefs. LGM, an analysis technique as yet rarely used to test mediation in applied studies, was found to be a viable way of testing for mediating factors. However temporality (timing of change) still could not be established as no assessments took place during the intervention period, so that while we know that change in fear-avoidance beliefs is related to change in disability, we cannot know *when* this change occurred, and whether change in fear-avoidance beliefs occurred first (as hypothesised) or whether in fact change in disability occurred before change in fear-avoidance beliefs.

Finally, in Chapter 8, exploratory analysis of clinical data collected as part of an Acceptance and Commitment Therapy (ACT) treatment programme for patients with chronic pain attending a pain clinic was conducted in order to investigate whether assessments made while an intervention is taking place might provide more information as to when factors change, to help improve claims for causality in studies of mediation. As in the BeBack analysis, this analysis was not mediation analysis but instead provided information on temporality, an important issue related to mediation analysis. Descriptive analyses were used to explore change at all time points (pre-, mid- and post-treatment) using both raw change and also different types of clinically useful change (improvement, large change, early or late change and consistent change). The different types of change were also entered into linear regression analyses to compare the predictive value of each type on change in disability outcome. Both group- and individual-level data were considered. It was found that change tended to occur later on in treatment, which was consistent with the ACT theoretical model and clinician expectations, but that this change was inconsistent and did not explain a significant proportion of change in disability. In terms of comparing pre-post assessment with the inclusion of mid-treatment assessments, the mid-treatment assessments were found to give more information as to when the variables of interest were likely to change and the individual-level data highlighted significant variability in scores. These preliminary findings provide some support for the calls to include mid-treatment assessments and to perform analysis which incorporates between- and within-participant change.

Overall, this thesis therefore makes a potentially important contribution to the literature on treatment mediation of psychological intervention studies for MSK pain through the identification of key design and analysis flaws in past treatment mediation research in this population and offers examples of mediation analysis using recommended design and analysis techniques. The aim of this final chapter is to outline guidance, based on the work presented here, on what steps

need to be taken to provide strong evidence of treatment targets in future psychologically oriented interventions for MSK pain, and specifically LBP. It will also highlight the limitations of the work carried out in the thesis and present its research and clinical implications.

9.2 Guidelines for Future Treatment Mediation Analysis

9.2.1 Theoretical basis for intervention

The importance of theoretical models lies in the framework they provide for understanding how an intervention leads to change in specific factors. The choice of theoretical model therefore depends on the factors of interest, or what the researchers think the intervention is targeting. In this thesis, the importance of basing the treatment on a theoretical model, and using this to generate hypotheses as to how the treatment will work, when to best assess key variables of interest and crucially, what factors to measure as mediators have all been highlighted. Several models of pain and disability (such as the fear-avoidance model (FAM), acceptance and commitment therapy (ACT) and misdirected problem solving model) were reviewed in Chapter 1, where it was highlighted that while links between the different model components are often tested, the model as a whole and whether the temporal order of the model components is correct, are not. The systematic review in Chapter 4 suggested that many psychological interventions for LBP are based somewhat loosely on psychological theory, often not declaring specific hypotheses as to exactly what factors are expected to change and making it difficult to compare between treatments because of a focus on different targets for change. Chapters 6, 7 and 8 in this thesis include analyses of studies where an *a priori* hypothesis as to how the intervention worked was included, although the ability to test this optimally was limited by different aspects of the study's designs. The Back In Action study and IMPACT treatment programme (Chapter's 7 and 8) in particular included measures that were closely linked with the

theoretical model being implemented (i.e. the factors that the model was trying to target). However, the lack of mid-treatment assessments in the Back In Action study and the lack of a control group in the IMPACT data means that caution is needed when interpreting the findings of these studies. Only the IMPACT data included in Chapter 8 explicitly tested any of the models outlined in Chapter 1 (ACT), which is consistent with the literature summarised in Chapter 1 that found few intervention studies explicitly based their intervention on a theoretical model. Despite this, the studies analysed highlight the importance of measuring the factors that are thought to be key in leading to change. Also, while most of the models outlined in Chapter 1 offer explanations for phenomena and how they develop (for example, the misdirected problem solving model hypothesises that persistent pain leads to worry and attempts to reduce pain, which if unsuccessful exacerbate the worry) they don't necessarily provide information on techniques to change such behaviours or thought processes. The current evidence to support such models rarely tests for mediating links between the variables in each model; for example, of the models summarised in Chapter 1, the FAM was the only model for which good evidence of mediating factors and analyses of the model as a whole could be found. However, there are examples of studies which have successfully developed a theoretically-based intervention. One study by Siemonsma et al (2010) demonstrated how, by designing a trial with clearly stated theoretical assumptions, we can better answer the question about how the trial works by including measures of the key factors expected to change and the outcome that is most relevant to measure. Another more recent example of an intervention designed using the Medical Research Council (MRC) guidelines for complex interventions, part of which advises identifying appropriate theory to guide intervention design, demonstrated that theory could be helpful in deciding on what factors to focus on and how to go about trying to change them (Carnes et al 2013). The key message is that identifying models led to the identification of specific factors to focus on trying to change, and specific techniques to change behaviours. Further work to identify what can be changed and how the change can be brought about is required to improve and test existing theoretical models.

It should be acknowledged that these studies, and many of the models outlined in Chapter 1, were developed and tested in secondary care populations, as were the tools commonly used to measure the different constructs. This may mean that they are unlikely to be suitable for primary care populations in which the influence of psychological factors such as fear-avoidance beliefs and depression may be less marked than in patients attending secondary or tertiary care, although the evidence presented in Chapters 6 and 7 indicated that strong mediating effects can still be found in primary care populations. The original STarT Back analysis outlined in Chapter 6 did start to address this by stratifying patients into groups based on prognostic factors, where patients who were found to have higher scores on psychological factors were treated with psychologically informed physiotherapy.

9.2.2 Validity and utility of the key psychometric measures

Pincus et al (2002) emphasise the importance of strong psychometric properties of tools used to measure psychological constructs and ensuring that the measures chosen are appropriate for use within a particular population. There has also been some question as to the distinctiveness of different psychological constructs in terms of available measurement tools. Strong correlations between factors may indicate they are not conceptually distinct (Sullivan et al 2001; Campbell et al 2013b) although this may be due to the items in the tool rather than the factor itself. Psychological factors are rarely examined in isolation, and are usually investigated alongside other potentially important factors, making it difficult to judge the importance of a single factor. In the STarT Back analysis (Chapter 6), a measure of psychological distress was created using measures of anxiety, depression, catastrophising thoughts and fear-avoidance beliefs, and in this analysis these measures were found to be distinct from measures of pain intensity, although a strong relationship between pain and distress was present. Most recently, a study by Campbell et al (2013b) carried out a detailed exploratory and confirmatory factor analysis on five psychological

scales commonly used in LBP research and found they could be accounted for primarily by four factors (Pain-related distress, Causal beliefs, Coping cognitions and Perceptions of the future). Pain-related distress in particular was found to be associated with pain and disability in their clinical sample of patients with LBP (*n*=1591) and accounted for much of the variance (15% and 28% for pain and disability respectively) while Causal beliefs and Perceptions of the future accounted for smaller proportions. The authors suggest that combining the various psychological constructs we routinely use in studies may make more useful markers for research, although this seems somewhat at odds with the recommendation of targeting specific factors seen elsewhere in the literature.

Overall then, when conducting mediation analysis assessments should be made of the psychometric properties of the measures to be used in the study to ensure they are appropriate for use in the population of interest and can adequately capture change in the constructs. This will help ensure that any mediating effects present are identified and can be attributed to a particular factor. If instruments have not been used on the population under investigation, additional validation or cross-validation should be undertaken. Should it not be possible to validate an instrument satisfactorily it may be necessary to develop a new tool to measure the construct in question (although this strategy may render comparison with other studies problematic).

9.2.3 Testing of potential mediator(s) and outcome during treatment

The assessment of key factors during an intervention can provide important information as to whether the assumed causal association between mediator and outcome is likely to be correct and also on the selection of the optimum assessment points for both mediator and outcome. Such an analysis may also provide information on when factors are likely to change, and can help better design assessment points in future mediation studies. This issue of temporality has been

discussed in the thesis but unfortunately none of the studies included in the secondary analyses or the systematic review offered an adequate evaluation of it. Even when measures were available during the treatment period (Chapter 8 – IMPACT data), the outcome of interest was not measured at the same time points as the potential mediators meaning that it was not possible to conclude that change in the mediators occurred prior to change in disability. However, it was clear from this analysis that assessments made during treatment do provide more information on how patients are responding to particular treatment targets, which is useful both clinically (as an audit or therapeutic tool) and for research purposes (to help decide at what time point to best measure a variable in order to pick up any change likely to occur).

9.3 Limitations of the Work Carried out for this Thesis

9.3.1 Missing data

As is to be expected in clinical studies of this type missing data has had to be accounted for in Chapters 5, 6, 7 and 8. This had to be dealt with in different ways for each analysis. In Chapters 5 (BeBACK) and 8 (IMPACT) all available data was used but response bias analysis was conducted to check for differences between responders and non-responders. In Chapter 6 (STarT Back), complete case analysis was used and in Chapter 7 (Back In Action) expectation maximisation (EM) was used to impute missing values. These datasets did have a large proportion of loss to follow-up, suggesting that a risk of bias was present. However the response bias analyses carried out on the data found few baseline differences between responders and non-responders on key variables of interest. In the STarT Back trial, the use of complete case analysis was due to the SEM software not being able to compute CIs for indirect effects and a specific goodness-of-fit index (SRMR) when the dataset contained missing data. A sensitivity analysis using all available data (but not providing confidence intervals) highlighted discrepancies in the data with responders to

the four-month follow-up point generally being 'healthier' at baseline (lower disability, catastrophising and fear-avoidance beliefs scores) compared to those who did not respond, suggesting that the results of the analysis are only representative of a specific group of patients. In the Back In Action study the analysis allowed imputation to be used more easily, and so this dataset was the only one where imputation was carried out.

Ideally, all efforts should be made to ensure data is not missing at the point of collection. While this was possible in the STarT Back and Back In Action trials and indeed far less missing data was present in these studies, the IMPACT data was collected for clinical rather than research purposes and therefore attempts to chase up any missing data were less stringent, and in the BeBACK study the baseline survey asked respondents to consent to further contact and only those who consented were sent a follow-up questionnaire, rather than all those who responded to the baseline survey. Where missing data is present and missing at random, imputation is the recommended way of dealing with missing values as it reduces the risk of attrition bias, but this can be complex to perform. This thesis highlights that regardless of the method used, conducting sensitivity analyses is important in helping to assess the impact of missing data.

9.3.2 Comparison of different mediation methods

The different methods of mediation analysis were chosen based on their suitability for the dataset and the hypothesised mediators. This resulted in different methods of mediation analysis being used for each analysis chapter, making comparison between the different mediation models difficult. While mediating effects were produced for both the STarT Back and Back In Action studies using the product of coefficients approach, the studies included different measures of mediator and outcome variables and different assessment points. In the STarT Back trial, it was hypothesised that clinicians targeted a broader concept of psychological distress rather than

targeting specific constructs individually (the training that clinicians delivering the intervention received focused on how to recognise and address each of the major targets used in pain management if and when they presented in a patient, but were not given a systematic programme to follow where each of the targets was addressed in turn). To account for this, SEM analysis, which allowed for the testing and inclusion of such a latent variable, was used. However this analysis cannot take into account more than two time points, so it could not be used to carry out analysis on the Back In Action trial where the purpose was to test the utility of including multiple assessments of mediator and outcome variables. The LGM used for the Back In Action data could not be used with the STarT Back data because it cannot include latent constructs, and SEM could not be used with the Back In Action data because it could not take into account the multiple measures of fear-avoidance beliefs and disability available. It was therefore not possible in this thesis to compare whether one method is superior to the other. This highlights the importance of designing a study *a priori* to look at mediating factors, and thinking carefully about how to design the study in order to conduct the strongest form of mediation analysis, by having all the necessary data available. The use of different methods however has served to illustrate the flexibility of mediation analysis and its ability to incorporate different research questions, variables and designs.

9.3.3 The measurement of change

The analysis of change has been central to this thesis. However, there is debate around what type of change to use, error associated with change measures, and how different types of change are interpreted and compared. In the analyses presented in this thesis both residualised (Chapters 5, 6 and 8) and raw change (Chapter 8) have been investigated, as well as growth trajectories (Chapter 7). While raw change is simple to understand and interpret, it does not account for baseline score which is often strongly predictive of outcome. Residualised change represents only

the change that occurred over and above baseline score, and numerous examples exist of residualised change scores in the applied literature of MSK pain research (e.g. George et al 2008; Bergbom et al 2012) that allow comparisons to be made across studies. However the values given by this change score are small in magnitude and more difficult to interpret, with some authors stating that this type of change is essentially representing error in the measure rather than any sort of change. Residualised change, like raw change, is also only between two assessment points whereas in this thesis it has been argued that analysis should include three or more assessment points in order to adequately capture change. Therefore, while residualised change scores have been found to be useful, future mediation studies should consider other, more sophisticated ways of analysing change.

Minimisation of measurement error is important in analysing change, especially in psychological research where much of the assessment of key factors is based on self-report and subjective interpretation. When measures are made over time, error is associated with each measurement (Streiner & Norman 2008). This is especially likely to be a problem in Chapter 8 (IMPACT data) where assessments are made frequently over a short period of time. Chapter 8 contains a discussion on the problems of the drawbacks of multiple assessment, including responder burden and increasing complexity of the subsequent analysis. However, it has been recommended that such measures are made frequently in order to adequately assess when factors change and in what order this change occurs. Full measures of constructs are not required at all time points; the analysis carried out in Chapter 8 utilised weekly data of individual items for each construct, and emerging evidence of similar repeated assessments using text messaging or electronic diaries have been shown to be useful in terms of collecting limited information on pain and mood (e.g. Karoly et al 2014), which could help alleviate responder burden. Statistical analysis techniques which allow for measurement error, such as SEM and LGM, have been shown in this thesis to be appropriate and valid methods to approach mediation analysis.

9.3.4 Direction of causality (temporality)

Key to mediation analysis is whether the hypothesised causal order of the variables can be shown to be correct. Each analysis chapter highlights different problems with establishing causality, although they progressively show more robust ways of looking at this issue. Chapter 8 provides an example of how temporality can possibly be investigated in future studies. However, all of the analyses presented are based on regression which makes strong assumptions about the causal links that can be made (see Chapter 3).

More recent advances in mediation analysis from epidemiology and statistics advocate a causal inference approach using the potential outcomes framework, which attempts to make these assumptions more explicit. In this framework a causal effect can be inferred from the difference between two potential outcomes; one that occurs if a person receives an intervention, and one that occurs if they do not receive an intervention (Egleston et al 2010; Imai et al 2010a). For any one participant, only one of these effects can ever be observed – if the participant had the intervention, we do not know what their outcome would have been if they had not had the intervention. The unobserved outcome is referred to as a counterfactual. Because only one outcome is ever observable, we can only ever know the average causal effect in an experimental study. The average treatment effect is also how treatment mediation is interpreted in regression and SEM models (the *a* and *c* paths) but this framework makes this more explicit. In the context of mediation analysis, this approach estimates a mediated effect by analysing how much the outcome would change in the intervention group if the mediator changed from the level it would take in the control arm to the level it would take in the intervention arm (Valeri 2013). This approach can also be extended to models where the outcome is not measured on an interval scale, which is currently difficult to do with regression-based models. However, it should be acknowledged that this framework is not necessarily better than the regression approaches;

advocates of this approach acknowledge that it still makes many of the assumptions of the regression approach and currently cannot take into account change over time as easily as techniques such as LGM. It should perhaps therefore be seen as another tool that can be applied to mediation questions depending on the study question, design, and available data.

9.4 Implications for the Design of Research

The research presented in this thesis also allows reflection on what steps need to be taken to improve the design of treatment mediation studies and clinical trials in the future. Four important issues have been identified. First, a key issue in mediation analysis, highlighted by the systematic review in Chapter 4, is the lack of power to identify a mediating effect due to small numbers of participants. Fritz & MacKinnon (2007)'s review of the sample sizes needed with different mediation methods emphasised the need to use appropriate methods that have the most power to detect an effect (if one exists) with the smallest possible sample size; the authors found that the bias-corrected bootstrap required the smallest sample size of all the mediation methods tested (maximum of n=462 depending on the size of the path coefficients). Bigger trials, or analysis that involves several trials through individual patient data (IPD) analysis, are therefore needed to help detect smaller mediating effects. Second, related to this, is that consistent use of measures for both potential mediators and outcome are needed to allow the combination of data and also to build consistent evidence. Third, the mediation analyses undertaken as part of this thesis highlighted the challenges around conducting mediation analysis as secondary analysis, and the need to pay specific attention in particular to careful measurement of potential mediators and outcomes in the design of the study. Finally, as was illustrated in Chapter 4, despite evidence being available on appropriate ways to test for mediating factors, sub-optimal analyses have been undertaken previously, and there is a clear need for sufficient statistical expertise in mediation

methods and analysis. While interest in mediation analysis is growing, leading to more evidence available in the field to learn from, a lack of appropriate statistical expertise in different mediation methods could lead to a lack of clarity in trial protocols and publications, and confusion over the interpretation of results. Given that this is a problem not just in the MSK pain field but also more widely within health psychology, it would seem important to encourage discussion and collaboration between the different fields to see what could be learnt from other areas.

Overall however, it is most important to recognise that the inclusion of mediation analysis is possible with just a few small changes to current trial design:

1. Consideration of theoretical models and mediation analysis methods in the very early stages of the trial development to allow identification of key mediating factors and crucially when these should be measured;
2. Cross-validation of the measurement tools of the constructs of interest to ensure they adequately represent the construct of interest and can capture change over time;
3. The inclusion of a plan of how the mediation analysis will be conducted.

Such considerations hopefully will help researchers to capture essential information on not just whether or not the intervention was effective but also why this was the case, enabling any future trials to build and improve on the results.

9.5 Clinical Implications

While specific comment on the nature of clinical interventions is beyond the remit of this thesis, the careful identification of treatment mediators has clinical as well as methodological implications. It might be anticipated that obtaining a clearer and more accurate evaluation of mediators will stimulate the development of more focused interventions, and provide

information on what parts of the intervention are key to changing outcome and when these changes occur. This in turn may help facilitate the identification of key targets for treatment, an essential prerequisite for the development of stratified care (Hill et al 2011; Foster et al 2014) now becoming a core feature of government policy. The identification of mediators (and moderators) might therefore be expected to be recognised as an important new direction in healthcare research.

9.6 Suggestions for a Further Research Agenda

As has been suggested, proper consideration of mediators will require a re-think of trial protocols, to include detail about mediating factors as early as possible. Thought needs to be given to what factors are likely to be treatment targets and how these will be influenced (a theoretical model is needed here to explain this process). Consideration of moderators (not a focus for this thesis) may help identify subgroups more likely to benefit from interventions. Moderated mediation analysis (e.g. Baron & Kenny 1986), or subgrouping patients in order to look at mediation in patients with specific characteristics, could help to see whether mediators could be more specifically targeted within these more homogenous groups. The STarT Back trial is one example of such stratified treatment, where patients were sub-grouped according to characteristics such as depression, fear-avoidance beliefs and comorbidity, but even when only people who scored highly on these factors were included in the analysis there was still considerable variability in terms of, for example, pain duration, age and treatment expectations. Investigating such heterogeneity further could perhaps identify more specific treatment targets. Moderated mediation analysis has not been discussed in this thesis but it is acknowledged that certain mediators may work better for patients with particular characteristics. However, there is a tension between identifying sub-groups for which specific mediating factors can be identified and

creating sub-groups so small that they are not practical for clinical or research use. Future research can start to address this by identifying broad subgroups based on a limited set of potential moderators and examining to what extent this improves the identification of mediating factors.

It should be recognised however that collection of additional information or more careful appraisal of mediators carries a cost burden, although hopefully this will be outweighed by healthcare savings resulting from improved treatment outcomes. Some funding bodies do already acknowledge the importance of investigating mediating factors (e.g. the National Institute for Health Research (NIHR) Efficacy and Mechanisms Evaluation Programme) but more stringent criteria need to be put in place to ensure only work which can adequately assess mediating factors is funded through such streams.

Three further research directions might be suggested. First, much of the research in treatment mediation in MSK pain studies focuses on factors known to be predictive of outcome but the work conducted in Chapter 5 of this thesis suggests that this does not necessarily mean these factors are good treatment targets and there is an opportunity for further work to be done in the validation of mediators. Qualitative approaches might be adopted to explore whether the factors often tested as mediators are truly mediators of outcome (i.e. are the factors that patients and clinicians see as key in leading to patient improvement) from both a patient and clinician perspective. This may help identify new potential mediators and give some evidence as to what factors are most important to test statistically as mediators. Second, the use of serial measurement, as in diary studies or single case designs, might provide a clearer picture of the temporal order and interdependence of key variables. Single case designs by definition include many repeated assessments of variables over time (Boersma et al 2004; Morley 2015), generating

detailed information on how the intervention worked for each individual participant (Flink et al 2015). Such designs could allow for specific mediating factors to be identified, and have already been shown to be useful in the identification of specific treatment pathways in the treatment of fear-avoidance (Vlaeyen et al 2001; Boersma et al 2004).

9.8 Conclusion

The overall aim of this thesis was to identify and evaluate the methods and analyses used in treatment mediation analysis in MSK pain intervention studies. This thesis has explored numerous methodological challenges in conducting treatment mediation analysis by identifying and critically appraising previous work in this area and providing applied examples using the current best available methods. This thesis has focused on issues around methodology because this is a significant first step in leading to the design of studies that can best test for mediating factors in clinical populations. The thesis provides examples of different methods of mediation analysis applied to primary and secondary care MSK pain populations and provides guidance for future work that will impact directly on clinical research and practice. The work presented here provides a start-point for new treatment mediation research, with mediation analysis being considered as a main study objective rather than as *post hoc* analysis. The recommendations from this thesis can be put into practice through simple changes to current efficacy trial design as a result of careful preliminary work to identify the factors most likely to lead to patient improvement. It is important that these changes are made in order for us to learn as much as possible from each intervention trial and therefore continually improve the care given to patients with MSK pain.

Reference List

Airaksinen, O., Brox, J.I., Cedraschi, C., Hildebrandt, J., Klaber-Moffett, J., Kovacs, F., Mannion, A.F., Reis, S., Staal, J.B., Ursin, H. & Zanoli, G. (2006). European guidelines for the management of chronic nonspecific low back pain. *European Spine Journal, **15 (Suppl 2)***, S192-S300.

Åkerblom, S., Perrin, S., Rivano Fischer, M. & McCracken, L.M. (2015). The mediating role of acceptance in multidisciplinary cognitive-behavioral therapy for chronic pain. *The Journal of Pain, **16 (7)***, 606-615.

Alcantara, M.A., Sampaio, R.F., Pereira, L.S.M., Fonseca, S.T., Silva, F.C.M., Kirkwood, R.N. & Mancini, M.C. (2010). Disability associated with pain – A clinical approximation of the mediating effect of beliefs and attitudes. *Physiotherapy Theory and Practice*, ***26 (7)***, 459-467.

Aldrich, S., Eccleston, C. & Crombez, G. (2000). Worrying about chronic pain: Vigilance to threat and misdirected problem solving. *Behaviour Research and Therapy, **38 (5)***, 457-470.

Anderson, J.C. & Gerbing, D.W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, **103 (3)***, 411-423.

Arnstein, P., Caudill, M., Mandle, C.L., Norris, A. & Beasley, R. (1999). Self efficacy as a mediator of the relationship between pain intensity, disability and depression in chronic pain patients. *Pain, **80 (3)**,* 483-491.

Arnstein, P. (2000). The mediation of disability by self efficacy in different samples of chronic pain patients. *Disability and Rehabilitation, **22 (17)**,* 794-801.

Balderson, B.H.K., Lin, E.H.B. & von Korff, M. (2004). The management of pain-related fear in primary care. In G.J.G. Asmundsen, J.W.S. Vlaeyen & G. Crombez (Eds.), *Understanding and treating fear of pain*. pp. 267-292. Oxford: Oxford University Press.

Ballinger, G.A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research methods, **7 (2)***, 127-150.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioural change. *Psychological Review, **84 (2)***, 191-215.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, **44 (9)***, 1175-1184.

Baranoff, J., Hanrahan, S.J., Kapur, D. & Conor, J.P. (2013). Acceptance as a process variable in relation to catastrophizing in multidisciplinary pain treatment. *European Journal of Pain, **17 (1)***, 101-110.

Baranoff, J., Hanrahan, S.J., Kapur, D. & Connor, J.P. (2014). Six month post-treatment deterioration in acceptance (CPAQ-8) and cognitions following multidisciplinary pain treatment. *Journal of Behavioural Medicine, **37 (3)***, 469-479.

Baranowski, T., Anderson, C & Carmack, C. (1998). Mediating variable framework in physical activity interventions. How are we doing? How might we do better? *American Journal of Preventive Medicine, **15 (4)***, 266-297.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, **42 (5)***, 815-824.

Barker, K.L., Reid, M. & Minns Lowe, C.J. (2009). Divided by a lack of common language? – A qualitative study exploring the use of language by health professionals treating back pain. *BMC Musculoskeletal Disorders, **10***, 123.

Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, **51 (6)***, 1173-1182.

Bekkering, G.E., Hendriks, H.J.M., van Tulder, M.W., Knol, D.L., Simmonds, M.J., Oostendorp, R.A.B. & Bouter, L.M. (2005). Prognostic factors for low back pain in patients referred for physiotherapy: Comparing outcomes and varying modelling techniques. *Spine, **30 (16)***, 1881-1886.

Bentler, P.M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, **42 (5)***, 825-829.

Bergbom, S., Boersma, K. & Linton, S.J. (2012). Both early and late changes in psychological variables relate to treatment outcome for musculoskeletal pain patients at risk of disability. *Behaviour Research and Therapy, **50 (11)***, 726-734.

Bergner, M., Bobbitt, R.A., Carter, W.B. & Gilson, B.S. (1981). The Sickness Impact Profile: Development and revision of a health status measure. *Medical Care, **19 (8)***, 787-805.

Bijsterbosch, J., Scharloo, M., Visser, A.W., Watt, I., Meulenbelt, I., Huizinga, T.W.J., Kaptein, A.A. & Kloppenburg, M. (2009). Illness perceptions in patients with osteoarthritis: Change over time and association with disability. *Arthritis & Rheumatism (Arthritis Care & Research),* ***61 (8)***, 1054-1061.

Bjelland, I., Dahl, A.A., Haug, T.T. & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research,* ***52 (2)***, 69-77.

Boersma, K., Linton, S., Overmeer, T., Jansson, M., Vlaeyen, J. & de Jong, J. (2004). Lowering fear-avoidance and enhancing function through exposure in vivo. A multiple baseline study across six patients with back pain. *Pain,* ***108 (1-2)***, 8-16.

Bolger, N., Davis, A & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology,* ***54***, 579-616.

Bombardier, C., Hayden, J. & Beaton, D.E. (2001). Minimally clinically important difference. Low back pain: outcome measures. *The Journal of Rheumatology,* ***28 (2)****,* 431-438.

Bränström, H. & Fahlström, M. (2008). Kinesiophobia in patients with chronic musculoskeletal pain: Differences between men and women. *Journal of Rehabilitation Medicine,* ***40 (5)***, 375-380.

Brown, R.L. (1997). Assessing specific mediational effects in complex theoretical models. *Schools of Medicine and Nursing Technical Report*, No. SON-2.96.

Brunner, E., De Herdt, A., Minguet, P., Baldew, S-S. & Probst, M. (2013). Can cognitive behavioural therapy based strategies be integrated into physiotherapy for the prevention of chronic low back pain? A systematic review. *Disability and Rehabilitation,* ***35 (1)***, 1-10.

Bullock, J.G., Green, D.P. & Ha, S.E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology,* ***98 (4)***, 550-558.

Burns, J.W., Kubilus, A., Bruehl, S., Harden, R.N. & Lofland, K. (2003). Do changes in cognitive factors influence outcome following multidisciplinary treatment for chronic pain? A cross-lagged panel analysis. *Journal of Consulting and Clinical Psychology,* ***71 (1)***, 81-91.

Burns, J.W., Nielson, W.R., Jensen, M.P., Heapy, A., Czlapinski, R. & Kerns, R.D. (2015). Specific and general therapeutic mechanisms in cognitive behavioral treatment of chronic pain. *Journal of Consulting and Clinical Psychology,* ***83 (1)***, 1-11.

Burton, A.K., Waddell, G., Tillotson, K.M. & Summerton, N. (1999). Information and advice to patients with back pain can have a positive effect: A randomized controlled trial of a novel educational booklet in primary care. *Spine,* ***24 (23)***, 2484-2491.

Byrne, B.M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment,* ***85 (1)***, 17-32.

Byrne, B.M., Lam, W.W.T. & Fielding, R. (2008). Measuring patterns of change in personality assessments: An annotated application of latent growth curve modelling. *Journal of Personality Assessment,* ***90 (6)***, 536-546.

Byrne, B.M. (2010). *Structural equation modelling with AMOS: Basic concepts, applications, and programming*. (2nd Ed.). London: Routledge.

Campbell, P., Foster, N.E., Thomas, E. & Dunn, K.M. (2013a). Prognostic indicators of low back pain in primary care: Five-year prospective study. *The Journal of Pain, **14 (8)***, 873-883.

Campbell, P., Bishop, A., Dunn, K.M., Main, C.J., Thomas, E. & Foster, N. (2013b). Conceptual overlap of psychological constructs in low back pain. *Pain, **154 (9)***, 1783-1791.

Carnes, S., Homer, K., Underwood, M., Pincus, T., Rahman, A. & Taylor, S.J.C. (2013). Pain management for chronic musculoskeletal conditions: The development of an evidence-based and theory-informed pain self-management course. *BMJ Open, **3***, e003534.

Cassidy, J.D., Côté, P., Carroll, L.J. & Kristman, V. (2005). Incidence and course of low back pain episodes in the general population. *Spine, **30 (24)**,* 2817-2823.

Cedraschi, C., Robert, J., Goerg, D. Perrin, E., Fischer, W. & Vischer, T.L. (1999). Is chronic non-specific low back pain chronic? Definitions of a problem and problems of a definition. *British Journal of General Practice, **49 (442)***, 358-362.

Cederaschi, C. & Allaz, A.F. (2005). How to identify patients with a poor prognosis in daily clinical practice. *Best Practice & Research Clinical Rheumatology, **19 (4)***, 577-591.

Cerin, E., Barnett, A. & Baranowski, T. (2009). Testing theories of dietary behavior change in youth using the mediating variable model with intervention programs. *Journal of Nutrition Education and Behaviour, **41 (5)***, 309-318.

Chen, H-T. (1990). *Theory-driven evaluations*. London: Sage.

Chen, H-T. (2015). *Practical program evaluation: Theory-driven evaluation and the integrated evaluation perspective* (2nd Ed.). London: Sage.

Cheong, J., MacKinnon, D.P. & Khoo, S.T. (2003). Investigation of mediational processes using parallel process latent growth curve modelling. *Structural Equation Modeling, **10 (2)***, 238.

Cheung, M.W-L. (2007). Comparison of methods of handling missing time-invariant covariates in latent growth models under the assumption of missing completely at random. *Organizational Research Methods, **10 (4)***, 609-634.

Cheung, G.W. & Lau, R.S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods, **11 (2)***, 296-325.

Chou, R. & Shekelle, P. (2010). Will this patient develop persistent disabling low back pain? *JAMA, **303 (13)***, 1295-1302.

Cleland, J., Gillani, R., Bienen, E.J. & Sadosky A. (2010). Assessing dimensionality and responsiveness of outcomes measures for patients with low back pain. *Pain Practice, **11 (1)***, 57-69.

Coffman, D.L. (2011). Estimating causal effects in mediation analysis using propensity scores. *Structural Equation Modeling, **18 (3)***, 357-369.

Cole, D.A. & Maxwell, S.E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, **112 (4)***, 558-577.

Collins, L.M., Graham, J.W. & Flaherty, B.P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research, **33 (2)***, 295-312.

Collins, L.M. & Graham, J.W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: Temporal design considerations. *Drug and Alcohol Dependence, **68 (Suppl 1)***, S85-S96.

Collins, L.M. (2006). Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology, **57***, 505-528.

Conway, J.M. & Huffcut, A.I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods, **6 (2)***, 147-168.

Cook, A.J., Brawer, P.A. & Vowles, K.E. (2006). The fear-avoidance model of chronic pain: Validation and age analysis using structural equation modelling. *Pain, **121 (3)***, 195-206.

Coryn, C.L.S., Noakes, L.A., Westine, C.D. & Schröter, D.C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, **32 (2)***, 199-226.

Costa, L. da C. M., Maher, C.G., McAuley, J.H., Hancock, M.J. & Smeets, R.J.E.M. (2011). Self-efficacy is more important than fear of movement in mediating the relationship between pain and disability in chronic low back pain. *European Journal of Pain*, ***15 (2)***, 213-219.

Costello, A.B. & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, **10 (7)***.

Coyne, J.C. & van Sonderen, E. (2012). No further research needed: Abandoning the Hospital Anxiety and Depression Scale (HADS). *Journal of Psychosomatic Research, **72 (3)***, 173-174.

Crombez, G., Eccleston, C., Van Damme, S., Vlaeyen, J.W.S. & Karoly, P. (2012). Fear-avoidance model of chronic pain: The next generation. *Clinical Journal of Pain, **28 (6)***, 475-483.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychmetrika, **16 (3),*** 297-334.

Curran, P.J. & Muthén, B.O. (1999). The application of latent curve analysis to testing developmental theories in intervention research. *American Journal of Community Psychology*, ***27 (4)***, 567-595.

Damsgård, E., Fors, T., Anke, A. & Røe, C. (2007). The Tampa Scale of Kinesiophobia: A rasch analysis of its properties in subjects with low back and more widespread pain. *Journal of Rehabilitation Medicine, **39 (9)***, 672-678.

Day, M.A. & Thorn, B.E. (2014). Using theoretical models to clarify shared and unique mechanisms in psychosocial pain treatments: A commentary on McCracken and Morley's theoretical paper. *The Journal of Pain, **15 (3)***, 237-238.

De Boer, M.J., Steinhagen, H.E., Versteegen, G.J., Strys, M.M.R.F. & Sanderman, R. (2014). Mindfulness, acceptance and catastrophising in chronic pain. *PLOS One, **9 (1)***, 1.

DeCoster, J (1998). *Overview of factor analysis*. Retrieved July 19th 2011 from http://www.stat-help.com/notes.html.

den Hollander, M., de Jong, J.R., Volders, S., Goossens, M.E.J.B., Smeets, R.J.E.M., & Vlaeyen, J.W.S. (2010). Fear reduction in patients with chronic pain: A learning theory perspective. *Expert Reviews Neurotherapeutics, **10 (11)***, 1733-1745.

Denison, E., Åsenlöf, P. & Lindberg, P. (2004). Self-efficacy, fear avoidance, and pain intensity as predictors of disability in subacute and chronic musculoskeletal pain patients in primary health care. *Pain, **111 (3)***, 245-252.

Dunn, K.M., Jordan, K. & Croft, P.R. (2003). Does questionnaire structure influence response in postal surveys? *Journal of Clinical Epidemiology, **56 (1)***, 10-16.

Dunn, K.M. & Croft, P.R. (2006). The importance of symptom duration in determining prognosis. *Pain, **121 (1-2)***, 126-132.

Dunn, K.M. & Cherkin, D.C. (2007). Letter to the editor: RMDQ-23. *Spine, **32 (2)***, 285-289.

Dunn, K.M., Croft, P.R., Main, C.J. & Von Korff, M. (2008). A prognostic approach to defining chronic pain: Replication in a UK primary care low back pain population. *Pain, **135 (1-2)***, 48-54.

Dunn, K.M., Campbell, P. & Jordan, K.P. (2013a). Long-term trajectories of back pain: Cohort study with 7-year follow-up. *BMJ Open*, ***3 (12)***, e003838.

Dunn, G., Emsley, R., Liu, H. & Landau, S. (2013b). Integrating biomarker information within trials to evaluate treatment mechanisms and efficacy for personalised medicine. *Clinical Trials, **10 (5)***, 709-719.

Eccleston, C. & Crombez, G. (1999). Pain demands attention: A cognitive-affective model of the interruptive function of pain. *Psychological Bulletin, **125 (3)**,* 356-366.

Eccleston, C. & Crombez, G. (2007). Worry and chronic pain: A misdirected problem solving model. *Pain, **132 (3)***, 233-236.

Eccleston, C., Williams, A.C.de C. & Morley, S. (2009). Psychological therapies for the management of chronic pain (excluding headache) in adults (review). Cochrane Database of Systematic Reviews, 2, No. CD007407.

Eccleston, C., Morley, S.J. & Williams, A.C. de C. (2013). Psychological approaches to chronic pain management: Evidence and challenges. *British Journal of Anaesthesia, **111 (1)***, 59-63.

Egleston, B.L., Cropsey, K.L., Lasev, A.B. & Heckman, C.J. (2010). A tutorial on principal stratification-based sensitivity analysis: Application to smoking cessation studies. *Clinical Trials, **7 (3)***, 286-298.

Ehde, D.M., Dillworth, T.M. & Turner, J.A. (2014). Cognitive-behavioural therapy for individuals with chronic pain: Efficacy, innovations, and directions for research. *American Psychologist, **69 (2)***, 153-166.

Emsley, R. & Dunn, G., & White, I.R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research, **19 (3)***, 237-270.

Enders, C.K. (2001a). The performance if the full estimation maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement, **61 (5)***, 713-740.

Enders, C.K. (2001b). A primer on maximum likelihood algorithms available for use with missing data. *Structural Equation Modeling, **8 (1)***, 128-141.

Enders, C.K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology, **56 (4)***, 267-288.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, **4 (3)***, 272-299.

Field, A. (2009). *Discovering statistics using SPSS* (3rd Ed.). London: Sage.

Fiedler, K., Schott, M. & Meiser, T. (2011). What mediation analysis can (not) do. *Journal of Experimental Social Psychology, **47 (6)***, 1231-1236.

Flink, I.K., Boersma, K., MacDonald, S. & Linton, S.J. (2012). Understanding catastrophizing from a misdirected problem-solving perspective. *British Journal of Health Psychology, **17 (2)***, 408-419.

Flink, I.K., Smeets, E., Bergbom, S. & Peters, M.L. (2015). Happy despite pain: Pilot study of a positive psychology intervention for patients with chronic pain. *Scandinavian Journal of Pain, **7 (1)***, 71-79.

Flor, H., Turk, D.C. & Birbaumer, N. (1985). Assessment of stress-related psychophysiological reactions in chronic back pain patients. *Journal of Consulting and Clinical Psychology, **53 (3)***, 354-364.

Focht, B.C., Rejeski, W.J., Ambrosius, W.T., Katula, J.A. & Messier, S.P. (2005). Exercise, self-efficacy, and mobility performance in overweight and obese older adults with knee osteoarthritis. *Arthritis & Rheumatism (Arthritis Care and Research), **53 (5)***, 659-665.

Fordyce, W.E. (1976). *Behavioral methods for chronic pain and illness*. Mosby: St Louis.

Forman, E.M., Chapman, J.E., Herbert, J.D., Goetter, E.M., Yuen, E.K. & Moitra, E. (2012). Using session-by-session measurement to compare mechanisms of action for acceptance and commitment therapy and cognitive therapy. *Behavior Therapy, **43 (2)***, 341-354.

Foster, N.E., Bishop, A., Thomas, E., Main, C., Horne, R., Weinman, J. & Hay, e. (2008). Illness perceptions of low back pain patients in primary care: What are they, do they change and are they associated with outcome? *Pain, **136 (1-2)***, 177-187.

Foster, N.E., Thomas, E., Bishop, A., Dunn, K.M. & Main, C.J. (2010). Distinctiveness of psychological obstacles to recovery in low back pain patients in primary care. *Pain, **148 (3)***, 398-406.

Foster, N.E., Mullis, R., Hill, J.C., Lewis, M., Whitehurst, D.G.T., Doyle, C., Konstantinou, K., Main, C., Somerville, S., Sowden, G., Wathall, S., Young, J., Hay, E.M. & IMPaCT Back Study Team. (2014). Effect of stratified care for low back pain in family practice (IMPaCT Back): A prospective population-based sequential comparison. *Annals of Family Medicine, **12 (2)***, 102-111.

Frazier, P.A., Tix, A.P. & Baron, K.E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, **51 (1)***, 115-134.

French, D.J., France, C.R., Vigneau, F., French, J.A. & Evans, R.T. (2007). Fear of movement/(re)injury in chronic pain: A psychometric assessment of the original English version of the Tampa scale for Kinesiophobia (TSK). *Pain, **127 (1-2)***, 42-51.

Fritz, M.S. & MacKinnon, D.P. (2007). Required sample size to detect the mediated effect. *Psychological Science, **18 (3)***, 233-239.

Fritz, M.S. Taylor, A.B. & MacKinnon, D.P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, **47 (1)***, 61-87.

Frost, H., Lamb, S.E., Doll, H.A., Taffe Carver, P. & Stewart-Brown, S. (2004). Randomised controlled trial of physiotherapy compared with advice for low back pain. *BMJ, **329 (7468)***, 708.

Frostholm, L., Oernboel, E., Christensen, K.S., Toft, T., Olsesen, F., Weinman, J. & Fink, P. (2007). Do illness perceptions predict health outcomes in primary care patients? A 2-year follow-up study. *Journal of Psychosomatic Research, **62 (2)***, 129-138.

Gaskin, C.J. & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies, **51 (3)***, 511-521.

Gatchel, R.J. & Turk, D.C. (2008). Criticisms of the biopsychosocial model in spine care: Creating and then attacking a straw person. *Spine, **33 (25)***, 2831-2836.

Gaudiano, B.A., Herbert, J.D. & Hayes, S.C. (2010). Is it the symptom or the relation to it? Investigating potential mediators of change in acceptance and commitment therapy for psychosis. *Behavior Therapy, **41 (4)***, 543-554.

Gauntlett-Gilbert, J., Connell, H., Clinch, J. & McCracken, L.M. (2013). Acceptance and values-based treatment of adolescents with chronic pain: Outcomes and their relationship to acceptance. *Journal of Pediatric Psychology, **38 (1)***, 72-81.

Geisser, M.E., Robinson, M.E., Miller, Q.L. & Bade, S.M. (2003). Psychosocial factors and functional capacity evaluation among persons with chronic pain. *Journal of Occupational Rehabilitation, **13 (4)***, 259-276.

Gelfand, L.A., Mensinger, J.L. & Tenhave, T. (2009). Mediation analysis: A retrospective snapshot of practice and more recent directions. *The Journal of General Psychology, **136 (2)***, 153-176.

George, S.Z., Zeppieri, G., Cere, A.L., Cere, M.R., Borut, M.S., Hodges, M.J., Reed, D.M., Valencia, C. & Robinson, M.E. (2008). A randomized trial of behavioral physical therapy interventions for acute and sub-acute low back pain (NCT00373867). *Pain, **140 (1)***, 145-157.

George, S.Z., Calley, D., Valencia, C. & Beneciuk, J.M. (2011). Clinical investigation of pain-related fear and pain catastrophizing for patients with low back pain. *Clinical Journal of Pain, **27 (2)***, 108-115.

Gheldof, E.L.M., Crombez, G., Van den Bussche, E., Vinck, J., Van Nieuwenhuyse, A., Moens, G., Mairiaux, P. & Vlaeyen, J.W.S. (2010). Pain-related fear predicts disability, but not pain severity: A path analytic approach of the fear-avoidance model. *European Journal of Pain, **14 (8)***, 870.e1-870.e9.

Gibbons, R.D., Hedeker, D. & DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology, **6***, 79-107.

Gillespie, D.F. & Streeter, C.L. (1994). Fitting regression models to research questions for analyzing change in nonexperimental research. *Social Work Research, **18 (4)***, 239-245.

Goubert, L., Crombez, G., Van Damme, S., Vlaeyen, J.W.S., Bijttebier, AP. & Roelofs, J. (2004). Confirmatory factor analysis of the Tampa Scale for Kinesiophobia. Invariant two-factor model across low back pain patients and fibromyalgia patients. *Clinical Journal of Pain, **20 (2)***, 103-110.

Greenland, S. & Rothman, K.J. (1998). Measures of effect and measures of association. In K.J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (2nd Ed.). London: Lippincott, Williams & Wilkins.

Grice, J.W., Cohn, A., Ramsey, R.R. & Chaney, J.M. (2015). On muddled reasoning and mediation modeling. *Basic and Applied Social Psychology, **37 (4)***, 214-225.

Grotle, M., Brox, J.I., Veierød, M.B., Glomsrød, B., Lønn, J.H. & Vøllestad, N.K. (2005). Clinical course and prognostic factors in acute low back pain. Patients consulting primary care for the first time. *Spine, **30 (8)***, 976-982.

Harland, N.J. & Georgieff, K. (2003). Development of the coping strategies questionnaire 24, a clinically utilitarian version of the coping strategies questionnaire. *Rehabilitation Psychology, **48 (4)***, 296-300.

Harland, N. & Martin, D. (2014). Exploring the longitudinal stability of the CSQ24 in a back pain population. *Rehabilitation Psychology, **59 (1)***, 79-84.

Hay, E.M., Mullis, R., Lewis, M., Vohora, K., Main, C.J., Watson, P., Dziedzic, K.S., Sim, J., Minns Lowe, C. & Croft, P.R. (2005). Comparison of physical treatments versus a brief pain-management programme for back pain in primary care: A randomised clinical trial in physiotherapy practice. *Lancet, **365 (9476)**,* 2024-2030.

Hay, E.M., Dunn, K.M., Hill, J.C., Lewis, M., Mason, E.E., Konstantinou, K., Sowden, G., Somerville, S., Vohora, K., Whitehurst, D. & Main, C.J. (2008). A randomised clinical trial of subgrouping and targeted treatment for low back pain compared with best current care. The STarT Back Trial study protocol. *BMC Musculoskeletal Disorders, **9***, 58.

Hayden, J.A., Chou, R., Hogg-Johnson, S. & Bombardier, C. (2009). Systematic reviews of low back pain prognosis had variable methods and results – guidance for future prognosis reviews. *Journal of Clinical Epidemiology, **62 (8)***, 781-796.

Hayden, J.A., Dunn, K.D., van der Windt, D.A. & Shaw, W.S. (2010). What is the prognosis of back pain? *Best Practice & Research Clinical Rheumatology, **24 (2)***, 167-179.

Hayduk, L. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement, **74 (6)***, 905-926.

Hayes, S.C., Luoma, J.B., Bond, F.W., Masuda, A. & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, **44 (1)***, 1-25.

Hayes, A.F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, **76 (4)***, 408-420.

Hayes, S.C., Levin, M.E., Plumb-Vilardaga, J., Villatte, J.L. & Pistorello, J. (2013). Acceptance and commitment therapy and contextual behavioural science: Examining the progress of a distinctive model of behavioural and cognitive therapy. *Behavior Therapy, **44 (2)***, 180-198.

Hayes, A.F. & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, **24 (10)***, 1918-1927.

Henschke, N., Kamper, S.J. & Maher, C.G. (2015). The epidemiology and economic consequences of pain. *Mayo Clinic Proceedings,* ***90 (1)***, 139-147.

Higgins, J.P.T. & Green, S. (2011). (Eds). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. Available from www.cochrane-handbook.org.

Hill, J.C. & Fritz, J.M. (2011). Psychosocial influences on low back pain, disability, and response to treatment. *Physical Therapy,* ***91 (5)***, 712-721.

Hill, S. (2010). The illness perceptions questionnaire – revised (IPQ-R). *Journal of Physiotherapy,* ***56 (4)***, 280.

Hill, J.C., Dunn, K.M., Lewis, M., Mullis, R., Main, C.J., Foster, N.E. & Hay, E.M. (2008). A primary care back pain screening tool: Identifying patient subgroups for initial treatment. *Arthritis & Rheumatism (Arthritis Care & Research)*, ***59 (5)***, 632-641.

Hill, J.C., Whitehurst, D.G., Lewis, M., Bryan, S., Dunn, K.M., Foster, N.E., Konstantinou, K., Main, C.J., Mason, E., Somerville, S., Sowden, G., Vohora, K. & Hay, E.M. (2011). Comparison of stratified primary care management for low back pain with current best practice (STarT Back): A randomised controlled trial. *Lancet,* ***378 (9802)***, 1560-1571.

Hobro, N., Weinman, J. & Hankins, M. (2004). Using the self-regulatory model to cluster chronic pain patients: The first step towards identifying relevant treatments? *Pain,* ***108 (3)***, 276-283.

Holbert, R.L. & Stephenson, M.T. (2002). Structural equation modeling in the communication sciences, 1995-2000. *Human Communication Research, **28 (4)***, 531-551.

Hong, J., Reed, C., Novick, D. & Happich, M. (2012). Costs associated with treatment of chronic low back pain: An analysis of the UK General Practice Research Database. *Spine, **38 (1)***, 75-82.

Hoy, D., March, L., Brooks, P., Blyth, F., Woolf, A., Bain, C., Williams, G., Smith, E., Vos, T., Barendregt, J., Murray, C., Burstein, R. & Buchbinder, R. (2014). The global burden of low back pain: Estimates from the Global Burden of Disease 2010 study. *Annals of the Rheumatic Diseases, **73 (6)***, 968-974.

Hoyle, R.H. & Robinson, J.C. (2004). Mediated and moderated effects in social psychological research. In C. Sansone, C.C. Morf & A.T. Panter (Eds.), *Handbook of Methods in Social Psychology* (pp.213-233). Thousand Oaks, CA: Sage.

Hoyt, W.T., Imel, Z.E. & Chan, F. (2008). Multiple regression and correlation techniques: Recent controversies and best practices. *Rehabilitation Psychology, **53 (3)***, 321-339.

Hu, L. & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity under parameterized model misspecification. *Psychological Methods, **3 (4)***, 424-453.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, **6 (1)***, 1-55.

Iacobucci, D., Saldanha, N. & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology, **17 (2)***, 139-153.

Imai, K., Keele, L. & Tingley, D. (2010a). A general approach to causal mediation analysis. *Psychological Methods, **15 (4)***, 309-334.

Imai, K., Keele, L. & Yamamoto, T. (2010b). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science, **25 (1)***, 51-71.

Imai, K. & Yamamoto, T. (2010). Causal inference with differential measurement error: Nonparametric identification and sensitivity analysis. *American Journal of Political Science, **54 (2)***, 543-560.

Imai, K., Jo, B. & Stuart, E.A. (2011). Using potential outcomes to understand causal mediation analysis: A comment on Maxwell, Cole & Mitchell (2011). *Multivariate Behavioural Research, **46 (5)***, 861-873.

Imai, K., Tingley, D. & Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society, Series A (Statistics in Society), **176 (1)***, 5-51.

Jackson, T., Wang, Y., Wang, Y. & Fan, H. (2014). Self-efficacy and chronic pain outcomes: A meta-analytic review. *The Journal of Pain, **15 (8)***, 800-814.

Jellema, P., van der Windt, D.A.W.M., van der Horst, H.E., Twisk, J.W.R., Stalman, W.A.B. & Bouter, L.M. (2005). Should treatment of (sub)acute low back pain be aimed at psychosocial prognostic factors? Cluster randomised clinical trial in general practice. *BMJ, **331 (7508)***, 84.

Johansson, P. & Høglend, P. (2007). Identifying mechanisms of change in psychotherapy: Mediators of treatment outcome. *Clinical Psychology and Psychotherapy, **14 (1)***, 1-9.

Johnson Wright, L., Zautra, A.J. & Going, S. (2008). Adaptation to early knee osteoarthritis: The role of risk, resilience, and disease severity on pain and physical functioning. *Annals of Behavioural Medicine*, ***36 (1)***, 70-80.

Jordan, K., Dunn, K.M., Lewis, M. & Croft, P. (2006). A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. *Journal of Clinical Epidemiology, **59 (1)***, 45-52.

Jordan, K.P., Kadam, U.T., Hayward, R., Porcheret, M., Young, C & Croft, P. (2010). Annual consultation prevalence of regional musculoskeletal problems in primary care: An observational study. *BMC Musculoskeletal Disorders, **11***, 144.

Judd, C.M. & Kenny, D.A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, **5 (5)***, 602-619.

Kahn, J.H. (2006). Factor analysis in counseling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, **34 (5)***, 684-718.

Kaiser, H.F. (1970). A second-generation little jiffy. *Psychometrika*, ***35 (4)***, 401-415.

Kaptein, A.A., Bijsterbosch,J., Scharloo, M., Hampson, S.E., Kroon, H.M. & Kloppenburg, M. (2010). Using the common sense model of illness perceptions to examine osteoarthritis change: A 6-year longitudinal study. *Health Psychology, **29 (1)***, 56-64.

Kamper, S.J., Maher, C.G., Hancock, M.J., Koes, K.W., Croft, P.R. & Hay, E. (2010). Treatment-based subgroups of low back pain: A guide to appraisal of research studies and a summary of current evidence. *Best Practice & Research Clinical Rheumatology, **24 (2)***, 181-191.

Karahalios, A., Baglietto, L., Carlin, J.B., English, D.R. & Simpson, J.A. (2012). A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. *BMC Medical Research Methodology, **12***, 96.

Karoly, P., Okun, M.A., Ruehlman, L.S. & Pugliese, J.A. (2008). The impact of goal cognition and pain severity on disability and depression in adults with chronic pain: An examination of direct effects and mediated effects via pain-induced fear. *Cognitive Therapy & Research*, ***32 (3)***, 418-433.

Karoly, P., Okun, M.A., Enders, C. & Tennen, H. (2014). Effects of pain intensity on goal schemas and goal persuit: A daily diary study. *Health Psychology, **33 (9)***, 968-976.

Kazdin, A.E. & Nock, M.K. (2003). Delineating mechanisms of change in child and adolescent therapy: Methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry, **44 (8)***, 1116-1129.

Kazdin, A.E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, **3***, 1-27.

Kazdin, A.D. (2009). Understanding how and why psychotherapy leads to change. *Psychotherapy Research, **19 (4-5)***, 418-428.

Keefe, F.J., Rumble, M.E., Scipio, C.E., Giordano, L.A. & Perri, L.C.M. (2004). Psychological aspects of persistent pain: Current state of the science. *The Journal of Pain, **5 (4)***, 195-211.

Keele, L. (2015). Causal mediation analysis: Warning! Assumptions ahead. *American Journal of Evaluation*, ***36 (4)***, 500-513.

Kelley, K. & Preacher, K.J. (2012). On effect size. *Psychological Methods, **17 (2)***, 137-152.

Kenny, D.A. (2008). Reflections on mediation. *Organizational Research Methods, **11 (2)**,* 353-358.

Kent, P.M. & Keating, J.L. (2008). Can we predict poor recovery from recent-onset nonspecific low back pain? A systematic review. *Manual Therapy, **13 (1)***, 12-28.

Kline, R.B. (2015). The mediation myth. *Basic and Applied Social Psychology, **37 (4)***, 202-213.

Koes, B.W., van Tulder, M.W. & Thomas, S. (2006). Diagnosis and treatment of low back pain. *BMJ, **332***, 1430-1434.

Koes, B.W., van Tulder, M., Lin, C-W.C., Macedo, L.G., McAuley, J. & Maher, C. (2010). An updated review of clinical guidelines for the management of non-specific low back pain in primary care. *European Spine Journal, **19 (12)***, 2075-2094.

Kole-Snijders, A.M.J., Vlaeyen, J.W.S., Goossens, M.E.J.B., Rutten-van Molken, M.P.M.H., Heuts, P.H.T.G., van Breukelen, G. & van Eek, H. (1999). Chronic low-back pain: What does cognitive coping skills training add to operant behavioural treatment? Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology, **67 (6)***, 931-944.

Kori, S.H., Miller, R.P. & Todd, D.D. (1990). Kinesiophobia: A new view of chronic pain behaviour. *Pain Management, **Jan/Feb***, 35-43.

Kosma, M., Ellis, R. & Bauer, J.J. (2012) Longitudinal changes in psychosocial constructs and physical activity among adults with physical disabilities. *Disability and Health Journal*, ***5 (1)***, 1-8.

Kraemer, H.C., Stice, E., Kazdin, A., Offord, D. & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, **158 (6)***, 848-856.

Kraemer, H.C., Wilson, G.T., Fairburn, C.G & Agras, W.S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, **59 (10)***, 877-883.

Kraemer, H.C., Kiernan, M., Essex, M. & Kupfer, D.J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology, **27, 2 (Suppl)***, S101-S108.

Laisné, F., Lecomte, C. & Corbière, M. (2012). Biopsychosocial predictors of prognosis in musculoskeletal disorders: A systematic review of the literature (corrected and republished). *Disability and Rehabilitation, **34 (22)***, 1912-1941.

Lakke, S.E., Soer, R., Takken, T. & Reneman, M.F. (2009). Risk and prognostic factors for non-specific musculoskeletal pain: A synthesis of evidence from systematic reviews classified into ICF dimensions. *Pain, **147 (1-3)***, 153-164.

Lamb, S.E., Lall, R., Hansen, Z., Castelnuovo, E., Withers, E.J., Nichols, V., Griffiths, F., Potter, R., Szczepura, A., Underwood, M., BeST trial group. (2010). A multicentered randomised controlled trial of a primary care-based cognitive behavioural programme for low back pain. The Back Skills Training (BeST) trial. *Health Technology Assessment, **14 (41)***, 1-253, iii-iv.

Lamé, I.E., Peters, M.L., Kessels, A.G., van Kleef, M. & Patijn, J. (2008). Test-retest stability of the Pain Catastrophizing Scale and the Tampa Scale for Kinesiophobia in chronic pain patients over a longer period of time. *Journal of Health Psychology, **13 (6)***, 820-826.

Laurenceau, J-P., Hayes, A.M. & Feldman, G.C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review, **27 (6)***, 682-695.

Lee, H., Hübscher, M., Moseley, G.L., Kamper, S.J., Traeger, A.C., Mansell, G. & McAuley, J.H. (2015). How does pain lead to disability? A systematic review and meta-analysis of mediation studies in people with back and neck pain. *Pain, **156 (6)***, 988-997.

Leeuw, M., Goossens, M.E.J.B., Linton, S.J., Crombez, G., Boersma, K. & Vlaeyen, J.W.S. (2007a). The fear-avoidance model of musculoskeletal pain: Current state of scientific evidence. *Journal of Behavioural Medicine*, ***30 (1)***, 77-94.

Leeuw, M., Houben, R.M.A., Severeijns, R., Picavet, H.S.J., Schouten, E.G.W. & Vlaeyen, J.W.S. (2007b). Pain-related fear in low back pain: A prospective study in the general population. *European Journal of Pain*, ***11 (3)***, 256-266.

Leeuw, M., Goossens, M.E.J.B., van Breukelen, G.J.P., de Jong, J.R., Heuts, P.H.T.G., Smeets, R.J.E.M., Köke, A.J.A. & Vlaeyen, J.W.S. (2008). Exposure in vivo versus operant graded activity in chronic low back pain patients: Results of a randomized controlled trial. *Pain, **138 (1)***, 192-207.

Leffondré, K., Abrahamowicz, M., Regeasse, A., Hawker, G.A., Badley, E.M., McCusker, J. & Belzile, E. (2004). Statistical measures were proposed for identifying longitudinal patterns of change in quantitative health indicators. *Journal of Clinical Epidemiology, **57 (10)***, 1049-1062.

Lemstra, M. & Olszynski, W.P. (2005). The effectiveness of multidisciplinary rehabilitation in the treatment of fibromyalgia. A randomized controlled trial. *Clinical Journal of Pain, **21 (2)***, 166-174.

Lethem, J., Slade, P.D., Troup, J.D.G. & Bentley, G. (1983). Outline of a fear-avoidance model of exaggerated pain perception – I. *Behaviour Research and Therapy, **21 (4)***, 401-408.

Leventhal, H., Brissette, I. & Leventhal, E.A. (2003). The common-sense model of self-regulation of health and illness. In L.D. Cameron & H. Leventhal (Eds.)., *The self-regulation of health and illness behaviour*, Chapter 3 pp 42-65. London: Routledge.

Li, T., Hutfless, S., Scharfstein, D.O., Daniels, M.J., Hogan, J.W., Little, R.J.A., Roy, J.A., Law, A.H., Dickersin, K. (2014). Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: A systematic review and expert consensus. *Journal of Clinical Epidemiology, **67 (1)***, 15-32.

Linton, S.J. (2000). A review of psychological risk factors in back and neck pain. *Spine, **25 (9)***, 1148-1156.

Linton, S.J. & Shaw, W.S. (2011). Impact of psychological factors in the experience of pain. *Physical Therapy, **91 (5),*** 700-711.

Little, R.J.A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, **83 (404)***, 1198-1202.

Liu, L.C., Flay, B.R. & Aban Aya Investigators. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya Youth Project drug prevention program. *Prevention Science, **10 (3)***, 197-207.

Lockhart, G., MacKinnon, D.P. & Ohlrich, V. (2011). Mediation analysis in psychosomatic medicine research. *Psychosomatic Medicine, **73 (1)***, 29-43.

Lubans, D.R., Foster, C. & Biddle, S.J.H. (2008). A review of mediators of behaviour in interventions to promote physical activity amongst children and adolescents. *Preventive Medicine, **47 (5),*** 463-470.

Lundberg, M.K.E., Styf, J. & Carlsson, S.G. (2004). A psychometric evaluation of the Tampa Scale for Kinesiophobia – from a physiotherapeutic perspective. *Physiotherapy Theory and Practice, **20 (2)***, 121-133.

Lundberg, M., Larsson, M., Östlund, H. & Styf, J. (2006). Kinesiophobia among patients with musculoskeletal pain in primary healthcare. *Journal of Rehabilitation Medicine, **38 (1)***, 37-43.

Lundberg, M., Grimby-Ekman, A., Verbunt, J. & Simmonds, M.J. (2011). Pain-related fear: A critical review of the related measures. *Pain Research and Treatment*, Article ID 494196.

Lynch, K.G., Cary, M., Gallop, R. & Ten Have, T.R. (2008). Causal mediation analyses for randomized trials. *Health Services and Outcomes Research Methodology*, ***8 (2)***, 57-76.

Macfarlane, G.J., Jones, G.T. & Hannaford, P.C. (2006). Managing low back pain presenting to primary care: Where do we go from here? *Pain, **122 (3)***, 219-222.

MacCallum, R.C. & Austin, J.T. (2000). Applications of structural equation modelling in psychological research. *Annual Review of Psychology, **51***, 201-226.

MacKinnon, D.P. & Dwyer, J.H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, **17 (2)***, 144-158.

MacKinnon, D.P. (1994). Analysis of mediating variables in prevention and intervention research. *National Institute on Drug Abuse Monograph, **139***, 127-153.

MacKinnon, D.P., Krull, J.L. & Lockwood, C.M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, **1 (4)***, 173-181.

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G. & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, **7 (1)***, 83-104.

MacKinnon, D.P., Lockwood, C.M. & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, **39 (1)***, 99-128.

MacKinnon, D.P., Fairchild, A.J. & Fritz, M.S. (2007). Mediation analysis. *Annual Review of Psychology, **58***, 593-614.

MacKinnon, D.P. (2008). *Introduction to statistical mediation analysis*. New York: Psychology Press.

MacKinnon, D.P. & Fairchild, A.J. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science, **18 (1)***, 16-20.

MacKinnon, D.P. (2011). Integrating mediators and moderators in research design. *Research on Social Work Practice, **21 (6)***, 675-681.

Main, C.J. & Watson, P.J. (1999). Psychological aspects of pain. *Manual Therapy, **4 (4)***, 203-215.

Main, C.J. & Williams, A.C.de C. (2002). Musculoskeletal pain. *BMJ, **325***, 534-537.

Main, C.J. & George, S.Z. (2011). Psychosocial influences on low back pain: Why should you care? *Physical Therapy, **91 (5)***, 609-613.

Main, C.J., Sowden, G. Hill, J.C., Watson, P.J. & Hay, E.M. (2012). Integrating physical and psychological approaches to treatment in low back pain: The development and content of the STarT Back trial's 'high-risk' intervention (STarT Back; ISRCTN37113406). *Physiotherapy,* ***98 (2),*** 110-116.

Main, C.J., Keefe, F.J., Jensen, M.P., Vlaeyen, J. & Vowles, K. (Eds). (2015). *Fordyce's behavioral methods for chronic pain and illness. Republished with commentaries*. Philadelphia: IASP Press (Wolters Kluwer Health).

Mallen, C.D., Peat, G., Thomas, E., Dunn, K.M. & Croft, P.R. (2007). Prognostic factors for musculoskeletal pain in primary care: A systematic review. *British Journal of General Practice,* ***57 (541)***, 655-661.

Maniadakis, N. & Gray, A. (2000). The economic burden of back pain in the UK. *Pain,* ***84 (1)***, 95-103.

Mansell, G., Kamper, S.J. & Kent, P. (2013). Why and how back pain interventions work: What can we do to find out? *Best Practice & Research Clinical Rheumatology,* ***27 (5)***, 685-697.

Mansell, G., Hill, J.C., Kamper, S.J., Kent, P., Main, C. & van der Windt, D.A. (2014). How can we design low back pain intervention studies to better explain the effects of treatment? *Spine,* ***39 (5)***, E305-E310.

March, L., Smith, E.U.R., Hoy, D.G., Cross, R.J., Sanchez-Riera, L., Blyth, F., Buchbinder, R., Vos, T. & Woolf, A. (2014). Burden of disability due to musculoskeletal (MSK) disorders. *Best Practice & Research Clinical Rheumatology,* ***28 (3)***, 353-366.

Maric, M., Wiers, R.W. & Prins, P.J.M. (2012). Ten ways to improve the use of statistical mediation analysis in the practice of child and adolescent treatment research. *Clinical Child and Family Psychology Review, **15 (3)***, 177-191.

Martin, A.L., Halket, E., Asmundsen, G.J.G., Flora, D.B. & Katz, J. (2010). Postraumatic stress symptoms and the diathesis-stress model of chronic pain and disability in patients undergoing major surgery. *Clinical Journal of Pain, **26 (6)***, 518-527.

Mathieu, J.E. & Taylor, S.R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behaviour. *Journal of Organizational Behavior, **27 (8)***, 1031-1056.

Mathieu, J.E., DeShon, R.P. & Bergh, D.D. (2008). Mediational inferences in organizational research. Then, now, and beyond. *Organizational Research Methods, **11 (2)***, 203-223.

Mayer, J.M., Haldeman, S., Tricco, A.C. & Dagenais, S. (2010). Management of chronic low back pain in active individuals. *Current Sports Medicine Reports, **9 (1)***, 60-66.

Maxwell, S.E. & Cole, D.A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, **12 (1)***, 23-44.

McAndrew, L.M., Musumeci-Szabó, T.J., Mora, P.A., Vileikyte, L., Burns, E., Halm, E.A., Levethal, E.A. & Leventhal, H. (2008). Using the common sense model to design interventions for the prevention and management of chronic illness threats: From description to process. *British Journal of Health Psychology, **13 (2)***, 195-204.

McCracken, L.M., Velleman, S.C. (2010). Psychological flexibility in adults with chronic pain: A study of acceptance, mindfulness, and values-based action in primary care. *Pain, **148 (1)***, 141-147.

McCracken, L.M. & Morley, S. (2014). The psychological flexibility model: A basis for integration and progress in psychological approaches to chronic pain management. *The Journal of Pain, **15 (3)***, 221-234.

McDonald, R.P. & Ho, M-H.R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, **7 (1)***, 64-82.

Menezes Costa, L. de C., Maher, C.G., Hancock, M.J., McAuley, J.H., Herbert, R.D. & Costa, L.O.P. (2012). The prognosis of acute and persistent low back pain: A meta-analysis. *Canadian Medical Association Journal (CMAJ), **184 (11)***, E613-E624.

Michie, S., Johnston, M., Francis, J., Hardeman, W. & Eccles, M. (2008). From theory to intervention: Mapping theoretically derived behavioural determinants to behaviour change techniques. *Applied Psychology: An International Review*, ***57 (4)***, 660-680.

Moore, J.E., Lorig, K., von Korff, M., Laurent, D. & Gonzalez, G. (1999).The back pain helpbook. Boston MA: Perseus Press.

Morley, S., Eccleston, C. & Williams, A. (1999). Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. *Pain, **80 (1-2),*** 1-13.

Morley, S. (2015). Single cases are complex. Illustrated by Flink et al. 'Happy despite pin: A pliot study of a positive psychology intervention for patients with chronic pain'. *Scandinavian Journal of Pain**, 7 (1)***, 55-57.

Morlion, B., Kempke, S., Luyten, P. , Coppens, E. & van Wambeke, P. (2011). Multidisciplinary pain education program (MPEP) for chronic pain patients: Preliminary evidence for effectiveness and mechanisms of change. *Current Medical Research & Opinion, **27 (8)**,* 1595-1601.

Moss-Morris, R., Weinman, J., Petrie, K.J., Horne, R., Cameron, L.D. & Buick, D. (2002). The revised illness perceptions questionnaire (IPQ-R). *Psychology and Health, **17 (1)***, 1-16.

Murphy, R., Cooper, Z., Hollon, S.D. & Fairburn, C.G. (2009). How do psychological treatments work? Investigating mediators of change. *Behaviour Research and Therapy, **47 (1)***, 1-5.

Mykletun, A., Stordal, E & Dahl, A. (2001). Hospital Anxiety and Depression (HAD) scale: Factor structure, item analyses and internal consistency in a large population. *British Journal of Psychiatry, **179***, 540-544.

Newman, D.A. (2003). Longitudinal modelling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, **6 (3)***, 328-362.

Nicassio, P.M., Radojevic, V., Weisman, M.H., Schuman, C., Kim, J. Schoenfeld-Smith, K. & Krall, T. (1997). A comparison of behavioral and educational interventions for fibromyalgia. *Journal of Rheumatology, **24 (10)***, 2000-2006.

NICE. Low back pain: Early management of persistent non-specific low back pain. National clinical practice guideline number CG88. London: NICE, 2009.

Nicholas, M.K. (2007). The pain self-efficacy questionnaire: Taking pain into account. *European Journal of Pain, **11 (2)***, 153-163.

Nock, M.K. (2007). Conceptual and design essentials for evaluating mechanisms of change. *Alcoholism: Clinical and Experimental Research. **31 (S3)***, 4S-12S.

Norton, P.J. & Asmundson, G.J.G. (2003). Amending the fear-avoidance model of chronic pain: What is the role of physiological arousal? *Behavior Therapy, **34***, 17-30.

O'Boyle, E.H. & Williams, L.J. (2011). Decomposing model fit: Measurement vs. theory in organizational research using latent variables. *Journal of Applied Psychology, **96 (1)***, 1-12.

Olinsky, A., Chen, S. & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modelling. *European Journal of Operational Research, **151 (1)***, 53-79.

Osman, A., Barrios, F.X., Gutierrez, P.M., Kopper, B.A., Merrifield, T. & Grittmann, L. (2000). The Pain Catastrophising Scale: Further psychometric evaluation with adult samples. *Journal of Behavioral Medicine, **23 (4)***, 351-365.

Ostelo, R.W.J.G. & de Vet, H.C.W. (2005). Clinically important outcomes in low back pain. *Best Practice & Research Clinical Rheumatology, **19 (4)***, 593-607.

Painter, J.E., Borba, C.P.C., Hynes, M., Mays, D. & Glanz, K. (2008). The use of theory in health behavior research from 2000 to 2005: A systematic review. *Annals of Behavioral Medicine, **35 (3)**,* 358-362.

Patrick, D.L., Deyo, R.A., Atlas, S.J., Singer, D.E., Chapin, A. & Keller, R.B. (1995). Assessing health-related quality of life in patients with sciatica. *Spine, **20 (17)***, 1899-1909.

Patient Education Media, Inc. (1996). Back pain (Recurrent). Videotape 1 (self-care). Time Life Medical: New York.

Pengel, L.H.M., Herbert, R.D., Maher, C.G. & Refshauge, K.M. (2003). Acute low back pain: Systematic review of its prognosis. *BMJ, **327 (7410)***, 323.

Philips, H.C. (1987). Avoidance behaviour and its role in sustaining chronic pain. *Behaviour Research and Therapy, **25 (4)***, 273-279.

Piehler, T.F., Bloomquist, M.L., August, G.J., Gewirtz, A.H., Lee, S.S. & Lee, W.S.C. (2014). Executive functioning as a mediator of conduct problems prevention in children of homeless families residing in temporary supportive housing: A parallel process latent growth modelling approach. *Journal of Abnormal Child Psychology, **42 (5)***, 681-692.

Pincus, T., Burton, A.K., Vogel, S. & Field, A.P. (2002). A systematic review of psychological factors as predictors of chronicity/disability in prospective cohorts of low back pain. *Spine, **27 (5)***, E109-E120.

Pincus, T., Vogel, S., Burton, A.K., Santos, R. & Field, A.P (2006). Fear-avoidance and prognosis in back pain: A systematic review and synthesis of current evidence. *Arthritis & Rheumatism,* ***54 (12)***, 3999-4010.

Pincus, T., Smeets, R.J.E.M., Simmonds, M.J. & Sullivan, M.J.L. (2010). The fear avoidance model disentangled: Improving the clinical utility of the fear avoidance model. *Clinical Journal of Pain*, ***26 (9)***, 739-746.

Pincus, T., Kent, P., Bronfort, G., Loisel, P., Pransky, G. & Hartvigsen, J. (2013). Twenty-five years with the biopsychosocial model of low back pain – Is it time to celebrate? A report from the twelfth international forum for primary care research on low back pain. *Spine,* ***38 (24)***, 2118-2123.

Powney, M., Williamson, P., Kirkham, J. & Kolamunnage-Dona, R. (2014). A review of the handling of missing longitudinal outcome data in clinical trials. *Trials,* ***15 (1)***, 237.

Preacher, K.J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology,* ***66***, 825-852.

Prestwich, A., Sniehotta, F.F., Whittington, C., Dombrowski, S.U., Rogers, L. & Michie, S. (2014). Does theory influence the effectiveness of health behaviour interventions? Meta-analysis. *Health Psychology,* ***33 (5)***, 465-474.

Quintana, S.M. & Maxwell, S.E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist,* ***27 (4)***, 485-527.

Ramond, A., Bouton, C., Richard, I., Roquelaure, Y., Baufreton, C., Legrand, E. & Huez, J-F. (2011). Psychosocial risk factors for chronic low back pain in primary care – A systematic review. *Family Practice, **28 (1)***, 12-21.

Ramond-Roquin, A., Bouton, C., Gobin-Tempereau, A-S., Airagnes, G., Richard, I., Roquelaure, Y. & Huez, J-F. (2014). Interventions focusing on psychosocial risk factors for patients with non-chronic low back pain in primary care – A systematic review. *Family Practice, **31 (4)**,* 379-388.

Rassen, J.A., Brookhart, A.M., Glynn, R.J., Mittleman, M.A. & Schneeweiss, S. (2009). Instrumental variables I: Instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *Journal of Clinical Epidemiology, **62 (12)***, 1226-1232.

Rhodes, R.E. & Pfaeffli, L.A. (2010). Mediators of physical activity behaviour change among adult non-clinical populations: A review update. *International Journal of Behavioural Nutrition and Physical Activity, **7***, 37.

Robinson, J.P., Theodore, B.R., Dansie, E.J., Wilson, H.D. & Turk, D.C. (2013). The role of fear of movement in subacute whiplash-associated disorders grades I and II. *Pain, **154 (3)***, 393-401.

Roe, R.A. (2012). What is wrong with mediators and moderators? *The European Health Psychologist, **14 (1)***, 4-10.

Roelofs, J., Goubert, L., Peters, M.L., Vlaeyen, J.W.S. & Crombez, G. (2004). The Tampa Scale for Kinesiophobia: Further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. *European Journal of Pain, **8 (5)***, 495-502.

Roelofs, J., Sluiter, J.K., Frings-Dresen, M.H.W., Goossens, M., Thibault, P., Boersma, K. & Vlaeyen, J.W.S. (2007). Fear of movement and (re)injury in chronic musculoskeletal pain: Evidence for an invariant two-factor model of the Tampa Scale for Kinesiophobia across pain diagnoses and Dutch, Swedish, and Canadian samples. *Pain, **131 (1-2)***, 181-190.

Roesch, S.C., Norman, G.J., Adams, M.A., Kerr, J., Sallis, J.F., Ryan, S., Calfas, K.J. & Patrick, K. (2009). Latent growth curve modelling of adolescent physical activity: Testing parallel process and mediation models. *Journal of Health Psychology, **14 (2)***, 313-325.

Roland, M. & Morris, R. (1983). A study of the natural history of back pain. Part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine, **8 (2)***, 141-144.

Roland, M. & Fairbank, J. (2000). The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine, **25 (24)***, 3115-3124.

Rosenstiel, A.K. & Keefe, F.J. (1983). The use of coping strategies in chronic low back pain patients: Relationship to patient characteristics and current adjustment. *Pain, **17 (1)***, 33-44.

Rothman, K.J. & Greenland, S. (1998). Causation and causal inference. In K.J. Rothman and S. Greenland (Eds.), *Modern epidemiology* (2nd Ed), London: Lippincott, Williams & Wilkins.

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. New York: J. Wiley & Sons.

Rucker, D.D., Preacher, K.J., Tormala, Z.L. & Petty, R.E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, **5 (6)***, 359-371.

Russell, D.W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin,* ***28 (12)***, 1629-1646.

Schafer, J.L. & Olsen, M.K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research,* ***33 (4)***, 545-571.

Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online,* ***8 (2)***, 23-74.

Schmitt, T.A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment,* ***29 (4)***, 304-321.

Schumacker, R.E. & Lomax, R.G. (2010). *A beginner's guide to structural equation modelling*. (3rd Ed.). London: Routledge.

Selig, J.P. & Preacher, K.J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development,* ***6 (2-3)***, 144-164.

Severeijns, R., van den Hout, M.A., Vlaeyen, J.W.S. & Picavet, J.S.J. (2002). Pain catastrophizing and general health status in a large Dutch community sample. *Pain,* ***99 (1-2)***, 367-376.

Seymour, R.B., Hughes, S.L., Campbell, R.T., Huber, G.M. & Desai, P. (2009). Comparison of two methods of conducting the Fit and Strong! Program. *Arthritis & Rheumatism (Arthritis Care and Research),* ***61 (7),*** 867-884.

Sharp, T.J. (2001). Chronic pain: A reformulation of the cognitive-behavioural model. *Behaviour Research and Therapy,* ***39 (7)***, 787-800.

Shelby, R.A., Somers, T.J., Keefe, F.J., Pells, J.J., Dixon, K.E. & Blumenthal, J.A. (2008). Domain specific self-efficacy mediates the impact of pain catastrophising on pain and disability in overweight and obese osteoarthritis patients. *The Journal of Pain*, ***9 (10)***, 912-919.

Shrout, P.E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods,* ***7 (4)***, 422-445.

Siemonsma, P.C., Schröder, C.D., Roorda, L.D. & Lettinga, A.T. (2010). Benefits of treatment theory in the design of explanatory trials: Cognitive treatment of illness perceptions in chronic low back pain rehabilitation as an illustrative example. *Journal of Rehabilitation Medicine,* ***42 (2)***, 111-116.

Siemonsma, P.C., Stuive, I., Roorda, L.D., Vollebregt, J.A., Lankhorst, G.J. & Lettinga, A.T. (2011). Best candidates for cognitive treatment of illness perceptions in chronic low back pain: Results of a theory-driven predictor study. *Journal of Rehabilitation Medicine,* ***43 (5)***, 454-460.

Sim, J. & Wright, C. (2000). *Research methods in health care*. Cheltenham: Stanley Thornes.

Sinharay, S., Stern, H.S. & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods,* ***6 (4)***, 317-329.

Skidmore, J.R., Koenig, A.L., Dyson, S.J., Kupper, A.E., Garner, M.J. & Keller, C.J. (2015). Pain self-efficacy mediates the relationship between depressive symptoms and pain severity. *Clinical Journal of Pain,* ***31 (2)***, 137-144.

Skinner, M., Wilson, H.D. & Turk, D.C. (2012). Cognitive-behavioral perspective and cognitive-behavioral therapy for people with chronic pain: Distinctions, outcomes, and innovations*. Journal of Cognitive Psychotherapy: An International Quarterly,* ***26 (2)***, 93-113.

Smeets, R.J.E.M., Vlaeyen, J.W.S., Kester, A.D.M. & Knotterus, J.A. (2006). Reduction of pain catastrophizing mediates the outcome of both physical and cognitive-behavioural treatment in chronic low back pain. *The Journal of Pain,* ***7 (4),*** 261-271.

Smeets, R., Köke, A., Lin, C-W., Ferreira, M. & Demoulin, C. (2011). Measures of function in low back pain/disorders. Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RDQ). *Arthritis Care & Research,* ***63 (S11)***, S158-S173.

Snelgrove, S., Edwards, S. & Liossi, C. (2013). A longitudinal study of patients' experiences of chronic low back pain using interpretative phenomenological analysis: Changes and consistencies. *Psychology & Health,* ***28 (2)***, 121-138.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* 1982 (pp. 290-312). Washington DC: American Sociological Association.

Söderlund, A. & Åsenlöf, P. (2010). The mediating role of self-efficacy expectations and fear of movement and (re)injury beliefs in two samples of acute pain. *Disability and Rehabilitation,* ***32 (25)***, 2118-2126.

Spencer, S.J., Zanna, M.P. & Fong, G.T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology,* ***89 (6)***, 845-851.

Spinhoven, P., ter Kuile, M., Kole-Snijders, A.M.J., Mansfeld, M.H., den Ouden, D-J. & Vlaeyen, J.W.S. (2004). Catastrophizing and internal pain control as mediators of outcome in the multidisciplinary treatment of chronic low back pain. *European Journal of Pain,* ***8 (3),*** 211-219.

Stanton, A.L., Luecken, L.J., MacKinnon, D.P. & Thompson, E.H. (2013). Mechanisms in psychosocial interventions for adults living with cancer: Opportunity for integration of theory, research, and practice*. Journal of Consulting and Clinical Psychology,* ***81 (2)***, 318-335.

Stephenson, M.T. & Holbert, R.L. (2003). A monte carlo simulation of observable versus latent variable structural equation modelling techniques. *Communication Research,* ***30 (3),*** 332-354.

Stone-Romero, E.F. & Rosopa, P.J. (2011). Experimental tests of mediation models: Prospects, problems, and some solutions. *Organizational Research Methods,* ***14 (4)***, 631-646.

Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency*. Journal of Personality Assessment,* ***80 (1)***, 99-103.

Streiner, D.L. & Norman, G.R. (2008). *Health measurement scales: A practical guide to their development and use.* (4th Ed). Oxford: Oxford University Press.

Sullivan, M.J.L., Bishop, S.R. & Pivik, J. (1995). The pain catastrophising scale: Development and validation. *Psychological Assessment,* ***7 (4)***, 524-532.

Sullivan, M.J.L., Thorn, B., Haythornthwaite, J.A., Keefe, F., Martin, M., Bradley, L.A. & Lefebvre, J.C. (2001). Theoretical perspectives on the relation between catastrophizing and pain. *Clinical Journal of Pain,* ***17 (1)***, 52-64.

Sullivan, M.J.L. & Simon, G. (2012). A telephonic intervention for promoting occupational re-integration in work-disabled individuals with musculoskeletal pain. *Translational Behavioral Medicine,* ***2 (2)***, 149-158.

Swinkels-Meewisse, E.J.C.M., Swinkels, R.A.H.M., Verbeek, A.L.M., Vlaeyen, J.W.S. & Oostendorp, R.A.B. (2003). Psychometric properties of the Tampa Scale for kinesiophobia and the fear avoidance beliefs questionnaire in acute low back pain. *Manual Therapy,* ***8 (1)***, 29-36.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (5th Ed.). Pearson.

Ten Have, T. & Joffe, M.M. (2010). A review of causal estimation of effects in mediation analysis. *Statistical Methods in Medical Research,* ***21 (1)***, 77-107.

Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M. & de Vet, H.C.W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology,* ***60 (1)***, 34-42.

Thoemmes, F., MacKinnon, D.P. & Reiser, M.R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling, **17 (3)***, 510-534.

Thomas, E., Silman, A.J., Croft, P.R., Papageorgiou, A.C., Jayson, M.I.V. & Macfarlane, G.J. (1999). Predicting who develops chronic low back pain in primary care: A prospective study. *BMJ, **318 (7199)***, 1662-1667.

Thorn, B.E. & Burns, J.W. (2011). Common and specific treatment mechanisms in psychosocial pain interventions: The need for a new research agenda. *Pain, **152 (4)***, 705-706.

Tinsley, H.E. & Tinsley, D.J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, **34 (4)**,* 414-424.

Tkachuk, G.A. & Harris, C.A. (2012). Psychometric properties of the Tampa Scale for Kinesiophobia-11 (TSK-11). *The Journal of Pain, **13 (10)***, 970-977.

Tomarken, A.J. & Waller, N.G. (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology, **112 (4)***, 578-598.

Tomarken, A.J. & Waller, N.G. (2005). Structural equation modelling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, **1***, 31-65.

Truchon, M., Côté, D., Filion, L., Arsenault, B. & Dionne, C. (2008). Low-back-pain related disability: An integration of psychological risk factors into the stress process model. *Pain, **137 (3)***, 564-573.

Turk, D.C. & Okifuji, A. (2002). Psychological factors in chronic pain: Evolution and revolution. *Journal of Consulting and Clinical Psychology, **70 (3)***, 678-690.

Turk, D.C. & Wilson, H.D. (2010). Fear of pain as a prognostic factor in chronic pain: Conceptual models, assessment, and treatment implications. *Current Pain and Headache Reports, **14 (2)***, 88-95.

Turner, J.A., Mancl, L. & Aaron, L.A. (2006). Short- and long-term efficacy of brief cognitive-behavioral therapy for patients with chronic temporomandibular disorder pain: A randomized, controlled trial. *Pain, **121 (3)***, 181-194.

Turner, J.A., Holtzman, S. & Mancl, L. (2007). Mediators, moderators, and predictors of therapeutic change in cognitive-behavioral therapy for chronic pain. *Pain, **127 (3)***, 276-286.

Twohig, M.P. (2012). Introduction: The basics of acceptance and commitment therapy. *Cognitive and Behavioral Practice, **19 (4)***, 499-507.

Ullman, J.B. (2007). Structural equation modelling. In B.G. Tabachnick & L.S. Fidell (Eds.), *Using multivariate statistics* (5th ed.), pp 676-780.

Valeri, L. & Vanderweele, T.J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods, **18 (2)***, 137-150.

van der Windt, D., Hay, E., Jellema, P. & Main, C. (2008). Psychosocial interventions for low back pain in primary care. Lessons learned from recent trials. *Spine, **33 (1),*** 81-89.

van Hartingsveld, F., Ostelo, R.W.J.G., Cuijpers, P., de Vos, R., Riphagen, I.I. & de Vet, H.C.W. (2010). Treatment-related and patient-related expectations of patients with musculoskeletal disorders. A systematic review of published measurement tools. *Clinical Journal of Pain,* ***26 (6)***, 470-488.

van Wilgen, C.P., van Ittersum, M.W. & Kaptein, A.A. (2013). Do illness perceptions of people with chronic low back pain differ from people without chronic low back pain? *Physiotherapy,* ***99 (1)***, 27-32.

Vergouw, D., Heymans, M.W., van der Windt, D.A.W.M., Foster, N.E., Dunn, K.M., van der Horst, H.E. & de Vet, H.C.W. (2012). Missing data and imputation: A practical illustration in a prognostic study on low back pain. *Journal of Manipulative and Physiological Therapeutics,* ***35 (6)***, 464-471.

Velicer, W.F. & Jackson, D.N. (1990). Component analysis versus common factor-analysis – Some further observations. *Multivariate Behavioral Research,* ***25 (1)***, 97-114.

Verkerk, K., Luijsterburg, P.A.J., Miedema, H.A., Pool-Goudzwaard, A. & Koes, B. (2012). Prognostic factors for recovery in chronic nonspecific low back pain: A systematic review. *Physical Therapy,* ***92 (9)***, 1093-1108.

Vlaeyen, J.W.S., Kole-Snijders, A.M.K., Boeren, R.G.B. & van Eek, H. (1995a). Fear of movement/(re)injury in chronic low back pain and its relation to behavioral performance. *Pain,* ***62 (3)***, 363-372.

Vlaeyen, J.W.S., Kole-Snijders, A.M.J., Rotteveel, A.M., Ruesink, R. & Heuts, P.H.T.G. (1995b). The role of fear of movement/(re)injury in pain disability. *Journal of Occupational Rehabilitation, **5 (4)***, 235-252.

Vlaeyen, J.W.S. & Linton, S.J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: A state of the art. *Pain, **85 (3)***, 317-332.

Vlaeyen, J.W.S, de Jong, J., Geilen, M., Heuts, P.H.T.G. & van Breukelen, G. (2001). Graded exposure in vivo in the treatment of pain-related fear: A replicated single-case experimental design in four patients with chronic low back pain. *Behaviour Research and Therapy, **39 (2)***, 151-166.

Vlaeyen, J.W.S. & Morley, S. (2005). Cognitive-behavioral treatments for chronic pain: What works for whom? *Clinical Journal of Pain, **21 (1)***, 1-8.

Vlaeyen, J.W.S., Crombez, G. & Linton, S.J. (2009). The fear-avoidance model of pain: We are not there yet. Comment on Wideman et al. “A prospective sequential analysis of the fear-avoidance model of pain” [Pain, 2009] and Nicholas “First things first: reduction in catastrophizing before fear of movement” [Pain, 2009]. *Pain, **146 (1-2),*** 222-223.

Von Korff, M., Balderson, B.H.K., Saunders, K., Miglioretti, D.L., Lin, E.H.B., Berry, S., Moore, J.E. & Turner, J.A. (2005). A trial of an activating intervention for chronic back pain in primary care and physical therapy settings. *Pain, **113 (3)***, 323-330.

Von Korff, M. & Miglioretti, D.L. (2005). A prognostic approach to defining chronic pain. *Pain, **117 (3)***, 304-313.

Von Korff, M. & Dunn, K.M. (2008). Chronic pain reconsidered. *Pain, **138 (2)***, 267-276.

Vowles, K.E., McCracken, L.M. & Eccleston, C. (2007). Processes of change in treatment for chronic pain: The contributions of pain, acceptance and catastrophizing. *European Journal of Pain, **11 (7)***, 779-787.

Vowles, K.E., McCracken, L.M. & Eccleston, C. (2008). Patient functioning and catastrophising in chronic pain: The mediating effects of acceptance. *Health Psychology*, ***27 (2)***, S136-S143.

Vowles, K.E., Witkiewitz, K., Sowden, G. & Ashworth, J. (2013). Acceptance and commitment therapy for chronic pain: Evidence of mediation and clinically significant change following an abbreviated interdisciplinary program of rehabilitation. *The Journal of Pain, **15 (1)***, 101-113.

Vowles, K.E., Sowden, G. & Ashworth, J. (2014). A comprehensive examination of the model underlying acceptance and commitment therapy for chronic pain. *Behavior Therapy, **45 (3)***, 390-401.

Waddell, G. (1987). A new clinical model for the treatment of low-back pain. *Spine, **12 (7)***, 632-644.

Wegener, S.T., Castillo, R.C., Haythornthwaite, J., MacKenzie, E.J., Bosse, M.J. & the LEAP Study Group. (2011). Psychological distress mediates the effect of pain on function. *Pain*, ***152 (6)***, 1349-1357.

Wertli, M.M., Rasmussen-Barr, E., Weiser, S., Bachmann, L.M. & Brunner, F. (2014a). The role of fear avoidance beliefs as a prognostic factor for outcome in patients with nonspecific low back pain: A systematic review. *The Spine Journal, **14 (5)***, 816-836.

Wertli, M.M., Eugster, R., Held, U., Steurer, J., Kofmehl, R. & Weiser, S. (2014b). Catastrophizing – A prognostic factor for outcome in patients with low back pain: A systematic review. *The Spine Journal, **14 (11)***, 2639-2657.

Wicksell, R.K., Olssen, G.L. & Hayes, S.C. (2010). Psychological flexibility as a mediator of improvement in acceptance and commitment therapy for patients with chronic pain following whiplash. *European Journal of Pain, **14 (10)***, 1059.e1-1059.e11.

Wicksell, R.K., Kemani, M., Jensen, K., Kosek, E., Kadetoff, D., Sorjonen, K., Ingvar, M. & Olsson, G.L. (2013). Acceptance and commitment therapy for fibromyalgia: A randomized controlled trial. *European Journal of Pain, **17 (4)***, 599-611.

Wideman, T.H., Adams, H. & Sullivan, M.J.L. (2009). A prospective sequential analysis of the fear-avoidance model of pain. *Pain, **145 (1-2)***, 45-51.

Wideman, T.H., Asmundson, G.G.J., Smeets, R.J.E.M., Zautra, A.J., Simmonds, M.J., Sullivan, M.J.L., Haythornthwaite, J.A. & Edwards, R.R. (2013). Rethinking the fear avoidance model: Toward a multidimensional framework of pain-related disability. *Pain, **154 (11)***, 2262-2265.

Williams, A. C. de C. (2002). Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences, **25 (4)***, 439-455.

Williams, J. & MacKinnon, D.P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling, **15 (1)***, 23-51.

Williams, L.J., Vandenberg, R.J. & Edwards, J.R. (2009). Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals, **3 (1)***, 543-604.

Witkiewitz, K., Donovan, D.M. & Hartzler, B, (2012). Drink refusal training as part of a combined behavioral intervention: Effectiveness and mechanisms of change. *Journal of Consulting and Clinical Psychology, **80 (3)***, 440-449.

Woby, S.R., Roach, N.K., Urmston, M. & Watson, P.J. (2005). Psychometric properties of the TSK-11: A shortened version of the Tampa Scale for Kinesiophobia. *Pain, **117 (1-2)***, 137-144.

Woby, S.R., Urmston, M. & Watson, P.J. (2007). Self-efficacy mediates the relation between pain-related fear and outcome in chronic low back pain patients. *European Journal of Pain, **11 (7)***, 711-718.

Wurm, L.H. & Fisicaro, S.A. (2014). What residualizing predictors in regression analyses does (and what it does *not* do). *Journal of Memory and Language, **72***, 37-48.

Wynne-Jones, G., Cowen, J., Jordan, J.L., Uthman, O., Main, C.J., Glozier, N. & van der Windt, D. (2014). Absence from work and return to work in people with back pain: A systematic review and meta-analysis. *Occupational and Environmental Medicine, **71 (6)***, 448-456.

Zale, E.L., Lange, K.L., Fields, S.A. & Ditre, J.W. (2013). The relation between pain-related fear and disability: A meta-analysis. *The Journal of Pain, **14 (10)***, 1019-1030.

Zhao, X.; Lynch Jr, J.G. & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis*. Journal of Consumer Research, **37 (2)***, 197-206.

Zigmond, A.S. & Snaith, R.P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, **67 (6)***, 361-370.

Appendix 4.1 Summary of changes made to critical appraisal tool

Lubans	**Cerin**	**Mansell**
Did the study cite a theoretical framework?	√	√
Were the study measures/procedures designed to influence mediating variables?	√	√
Were pilot studies conducted/reported to test the effect of the intervention on mediators?	*X*	√
Was an objective measure of physical activity used?	*X*	*X*
Were the psychometric characteristics of mediator variables reported and were they within accepted ranges (Cronbach's alpha and test-retest reliability >.60)?	√	√
Did the study report a power calculation and was the study adequately powered to detect mediation?	√	√
Did the study use an experimental design?	√	√
Was post-intervention physical activity controlled for baseline physical activity?	*X*	*X*
	Were all 3 steps for testing validity of a theory of behaviour change performed?	*X*
	Were the psychometric characteristics of the outcome measure reported, and were they within acceptable ranges (i.e. test-retest or Cronbach's alpha >.60?)	√
	Were statistically appropriate/acceptable methods of data analysis used?	√
	Did the study ascertain whether changes in the mediating variables preceded changes in the outcome variables?	√
		Did the study report a change between baseline and follow-up for each mediator tested/reported?
		Was the change in the potential mediator correlated with change in outcome?
		Did the study control for possible confounding factors, e.g. baseline values?

Appendix 4.2 Template of data extraction form and critical appraisal tool

Systematic review of mediation studies in MSK pain – Data extraction and critical appraisal form

1st Author & Year Published: ..... Reviewer Initials: .....

Country.....

Type of study: RCT ☐

Before & After ☐

Other (please specify) ☐ .....

Unclear ☐

Study Setting: Primary Care ☐

Secondary Care ☐

Tertiary Care ☐

Other (please specify) ☐ .....

Unclear ☐

Patients suffering from:		Chronic	Sub-acute	Acute
MSK Pain	☐	☐	☐	☐
Chronic Pain (Unspecified)	☐	☐	☐	☐

Pain region (if given)……..

Study Aims/Hypotheses:

1. ……
   ……
   ……
   ……
2. ……
   ……
   ……
   ……
3. ……
   ……
   ……
   …….

Critical Appraisal Item 1	Yes	No	Unclear/Not reported
Did the study cite a theoretical framework?	☐	☐	☐
Critical Appraisal Item 2	Yes	No	Unclear/Not reported
Did the study use an experimental design?	☐	☐	☐
Critical Appraisal Item 3	Yes	No	Unclear/Not reported
Were pilot studies reported/conducted to test the effect of the intervention on mediators?	☐	☐	☐

Inclusion Criteria	
Exclusion Criteria	

Intervention Group: *What happened to participants in this group?*

- Number of participants……..
- Number who dropped out……..
- Who conducted the intervention……
- Duration of intervention……
- Number of sessions…….

What did the intervention involve?

Control / Comparator Group: *What happened to participants in this group?*

- Number of participants……..
- Number who dropped out……..
- Who conducted the control/comparator…….
- Duration of control/comparator……..
- Number of sessions……..

What did the control/comparator involve?

Length of follow-up……

Follow-up points……..

Outcome

Outcome	Measure used (Full name and Acronym)	Reference and Details of Measure

Critical Appraisal Item 4 reported

	Yes	No	Unclear/Not
Were the psychometric characteristics of the outcome measure reported, and were they within acceptable ranges (i.e. test-retest reliability or Cronbach's α >.60?)	☐	☐	☐

Mediation

Mediator	Measure used (Full name and Acronym)	Reference and Details of Measure

Critical Appraisal Item 5 reported

	Yes	No	Unclear/Not
Were the psychometric characteristics of the mediator variables reported, and were they within acceptable ranges?	☐	☐	☐

Method(s) of mediation analysis used.....

	Yes	No	Unclear/Not reported
Was bootstrapping conducted?	☐	☐	☐

Details of bootstrapping……

	Yes	No	Unclear/Not reported
Are CIs/effect sizes given?	☐	☐	☐

Details……

Preliminary analysis conducted……
……
……

	Yes	No	Unclear/Not reported
Were patient's scores converted to change scores for analysis?	☐	☐	☐

Details of change scores……
……

Critical Appraisal Item 6	Yes	No	Unclear/Not Reported
Did the study report change between baseline and follow-up for the potential mediator(s)?	☐	☐	☐

Critical Appraisal Item 7	Yes	No	Unclear/Not Reported
Was the change in the potential mediator(s) correlated with change in outcome?	☐	☐	☐

Critical Appraisal Item 8	Yes	No	Unclear/Not Reported
Were the study methods/procedures designed to influence mediating variables?	☐	☐	☐

Critical Appraisal Item 9	Yes	No	Unclear/Not Reported
Did the study report a power calculation, and was the study powered to detect mediation?	☐	☐	☐

Critical Appraisal Item 10	Yes	No	Unclear/Not Reported
Were statistically appropriate/acceptable methods of data analysis used?	☐	☐	☐

Critical Appraisal Item 11	Yes	No	Unclear/Not Reported
Did the study control for possible confounding factors, e.g. baseline values?	☐	☐	☐

Details……

Critical Appraisal Item 12	Yes	No	Unclear/Not Reported
Did the study ascertain whether changes in the mediating variable(s) precede changes in the outcome variable(s)?	☐	☐	☐

Findings

Which of the tested variables were found to mediate outcome(s)? *(If more than one outcome, please state which outcome you are referring to*)

……
……
……

Please add further comments/information relating to this study if necessary:

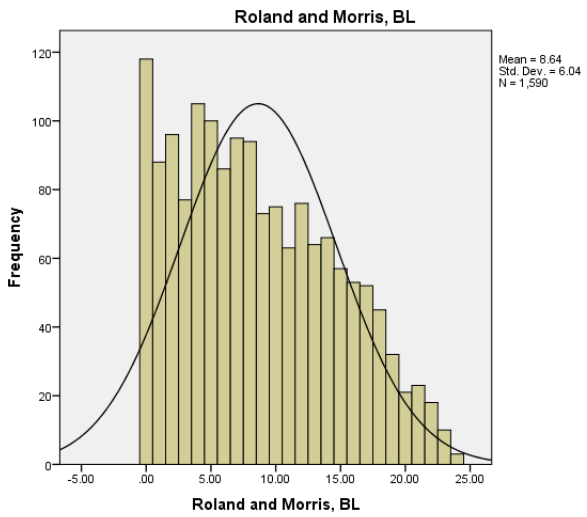
……
……
……
……

Appendix 5.1 Skewness and kurtosis values for the variables included in the BeBACK secondary analysis

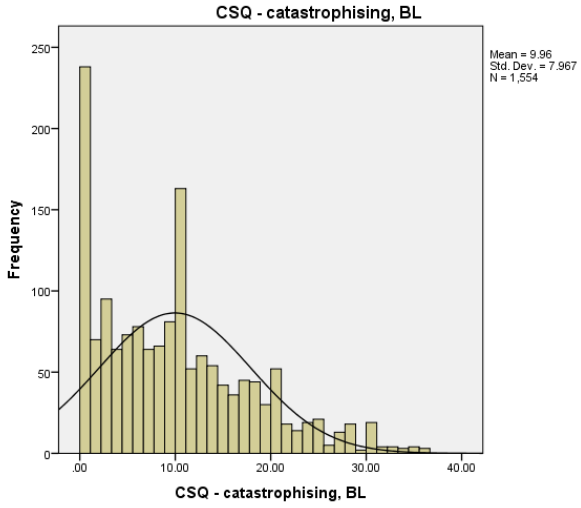
Baseline and three-, six- and 12-month follow-up

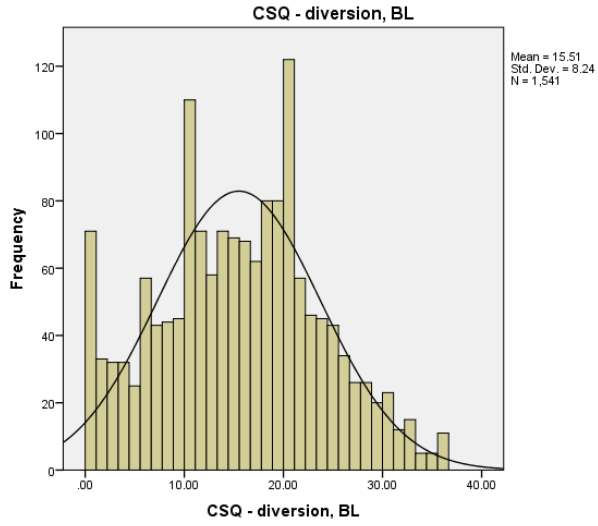
Measure		**Baseline (Mean (SD)) (*n*=1591)**	**Skewness (SE)**	**Kurtosis (SE)**	**3-month follow-up (Mean (SD)) (*n*=856)**	**Skewness (SE)**	**Kurtosis (SE)**	**6-month follow-up (Mean (SD)) (*n*=810)**	**Skewness (SE)**	**Kurtosis (SE)**	**12-month follow-up (Mean (SD)) (*n*=473)**	**Skewness (SE)**	**Kurtosis (SE)**
RMDQ		8.64 (6.04)	0.40 (0.06)	-0.80 (0.12)	6.61 (6.09)	0.87 (0.08)	-0.25 (0.17)	6.19 (6.14)	0.98 (0.09)	-0.09 (0.17)	5.83 (6.12)	1.02 (0.11)	0.03 (0.22)
Self-efficacy		37.81 (14.56)	-0.38 0.06)	-0.69 (0.12)	39.96 (14.11)	-0.57 (0.09)	-0.46 (0.19)	40.80 (14.15)	0.02 (0.10)	0.90 (0.20)	42.17 (14.07)	-0.86 (0.13)	0.11 (0.26)
Coping	Catastrophising	9.96 (7.97)	0.80 (0.06)	0.14 (0.12)	8.50 (7.95)	1.15 (0.09)	0.97 (0.19)	8.70 (7.51)	1.00 (0.10)	0.74 (0.20)	8.19 (7.20)	1.00 (0.13)	0.67 (0.26)
	Diversion	15.51 (8.24)	0.08 (0.06)	-0.53 (0.13)	***	***	***	15.49 (8.71)	0.07 (0.10)	-0.64 (0.20)	14.68 (8.42)	0.12 (0.13)	0.77 (0.26)
	Re-interpretation	7.83 (6.99)	0.94 (0.06)	0.46 (0.12)	***	***	***	8.71 (7.29)	0.86 (0.10)	0.18 (0.20)	6.83 (7.00)	0.90 (0.13)	0.77 (0.26)
	Cognitive coping	16.27 (6.48)	-0.13 (0.06)	-0.37 (0.12)	***	***	***	17.29 (6.28)	-0.12 (0.10)	-0.49 (0.20)	17.74 (6.26)	-0.30 (0.13)	-0.28 (0.26)
Fear-avoidance beliefs		39.67 (6.89)	0.23 (0.06)	0.77 (0.13)	***	***	***	38.38 (6.30)	-0.04 (0.10)	0.59 (0.20)	37.80 (6.11)	0.05 (0.13)	0.75 (0.26)
Anxiety		8.26 (4.55)	0.30 (0.06)	-0.47 (0.12)	***	***	***	6.60 (4.57)	-0.54 (0.09)	-0.23 (0.18)	6.31 (4.48)	0.56 (0.11)	-0.32 (0.23)
Depression		6.53 (4.37)	0.61 (0.06)	-0.18 (0.12)	***	***	***	5.08 (4.28)	0.82 (0.09)	0.10 (0.18)	4.67 (4.32)	0.98 (0.11)	0.32 (0.23)
Illness perceptions	Personal control	20.51 (3.79)	-0.10 (0.06)	-0.04 (0.12)	20.77 (4.17)	-0.26 (0.08)	0.42 (0.17)	20.93 (4.16)	-0.15 (0.09)	-0.05 (0.18)	21.26 (4.27)	-0.37 (0.11)	0.45 (0.23)

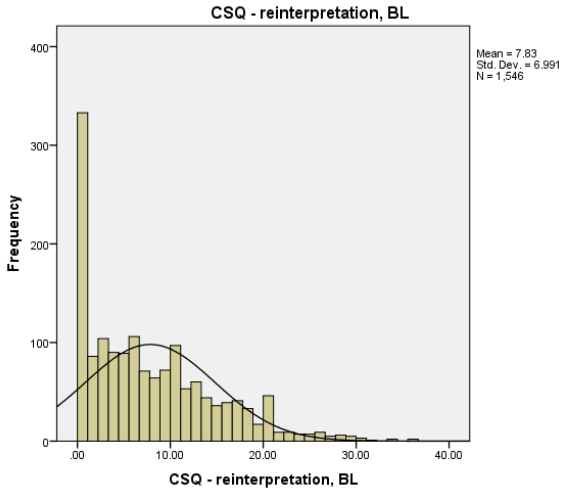
	Acute-chronic timeline	19.64 (5.84)	-0.17 (0.06)	-0.68 (0.12)	20.92 (5.81)	-0.57 (0.08)	-0.14 (0.17)	21.00 (5.80)	-0.59 (0.09)	-0.14 (0.18)	21.38 (5.91)	-0.73 (0.12)	0.07 (0.23)
	Identity	4.03 (2.37)	0.60 (0.06)	0.91 (0.12)	***	***	***	3.14 (2.41)	0.23 (0.09)	2.13 (0.18)	3.02 (2.23)	0.58 (0.11)	2.34 (0.23)
	Consequences	17.33 (5.48)	0.07 (0.06)	-0.69 (0.12)	17.06 (5.82)	0.09 (0.08)	-0.63 (0.17)	16.49 (5.80)	0.15 (0.09)	-0.73 (0.18)	16.07 (5.88)	0.28 (0.11)	-0.64 (0.23)
	Timeline cyclical	13.04 (3.39)	-0.46 (0.06)	-0.15 (0.12)	13.02 (3.40)	-0.61 (0.08)	0.07 (0.17)	12.82 (3.52)	-0.59 (0.09)	-0.07 (0.18)	12.85 (3.31)	-0.62 (0.12)	0.33 (0.23)
	Illness coherence	13.77 (4.99)	0.23 (0.06)	-0.73 (0.12)	12.72 (4.91)	0.53 (0.08)	0.07 (0.17)	11.88 (4.61)	0.59 (0.09)	-0.05 (0.18)	11.41 (4.49)	0.73 (0.11)	0.19 (0.23)
	Treatment control	16.99 (3.36)	-0.36 (0.06)	0.37 (0.12)	16.51 (3.52)	-0.45 (0.09)	0.60 (0.17)	16.58 (3.62)	-0.27 (0.09)	0.52 (0.18)	16.68 (3.55)	-0.28 (0.11)	0.76 (0.23)
	Emotional representation	16.71 (5.23)	0.17 (0.06)	-0.42 (0.12)	16.52 (5.20)	0.14 (0.08)	-0.24 (0.17)	15.69 (5.29)	0.21 (0.09)	-0.35 (0.18)	15.23 (5.18)	0.26 (0.11)	-0.32 (0.23)
	Psychological attribution	11.93 (4.11)	0.42 (0.06)	0.23 (0.12)	***	***	***	11.94 (4.25)	0.50 (0.09)	0.17 (0.18)	11.86 (3.94)	0.29 (0.12)	-0.22 (0.23)
	Risk factors	15.02 (4.15)	0.14 (0.06)	0.48 (0.12)	***	***	***	15.42 (3.95)	0.05 (0.09)	0.53 (0.18)	15.53 (3.86)	0.02 (0.12)	0.31 (0.23)
	Immunity	5.33 (1.95)	0.43 (0.06)	-0.22 (0.12)	***	***	***	5.29 (1.96)	0.47 (0.09)	-0.00 (0.18)	5.17 (1.86)	0.58 (0.12)	0.06 (0.23)
	Accident/Chance	5.98 (1.92)	-0.13 (0.06)	-0.30 (0.13)	***	***	***	5.86 (1.92)	-0.09 (0.09)	-0.29 (0.18)	6.00 (2.01)	-0.05 (0.12)	-0.42 (0.24)

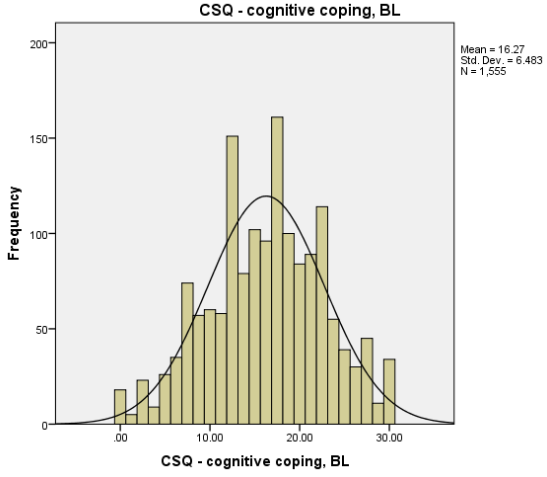

Roland and Morris, BL
Mean = 8.64
Std. Dev. = 6.04
N = 1,590
Frequency
Roland and Morris, BL

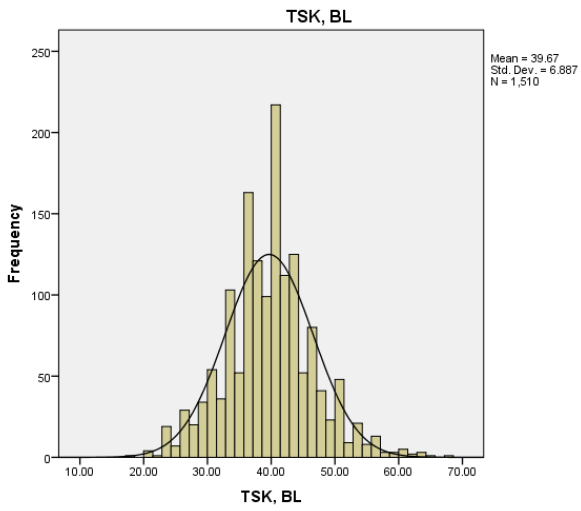

Self efficacy score, BL
Mean = 37.81
Std. Dev. = 14.565
N = 1,567
Frequency
Self efficacy score, BL

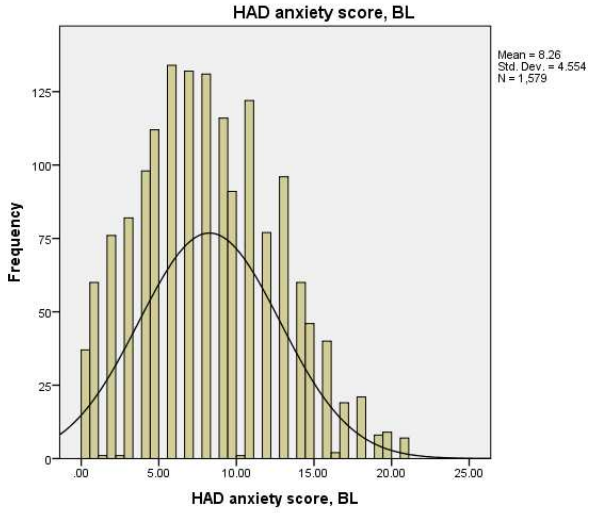

CSQ - catastrophising, BL
Mean = 9.96
Std. Dev. = 7.967
N = 1,554
Frequency
CSQ - catastrophising, BL

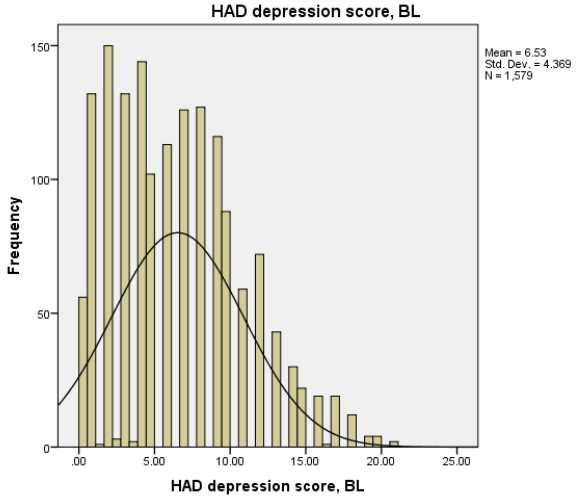

CSQ - diversion, BL
Mean = 15.51
Std. Dev. = 8.24
N = 1,541
Frequency
CSQ - diversion, BL

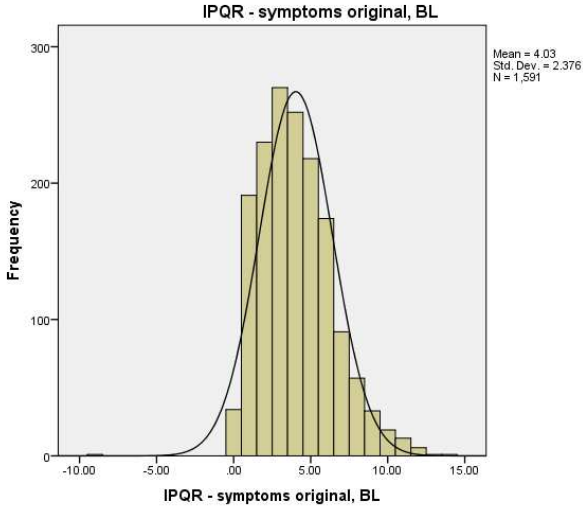

CSQ - reinterpretation, BL
Mean = 7.83
Std. Dev. = 6.991
N = 1,546
Frequency
CSQ - reinterpretation, BL

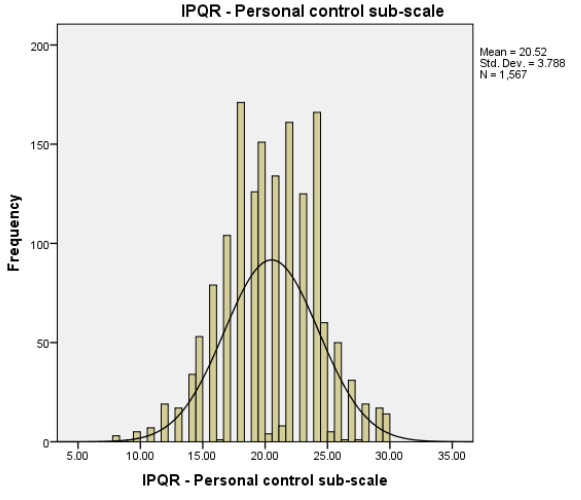

CSQ - cognitive coping, BL
Mean = 16.27
Std. Dev. = 6.483
N = 1,555
Frequency
CSQ - cognitive coping, BL

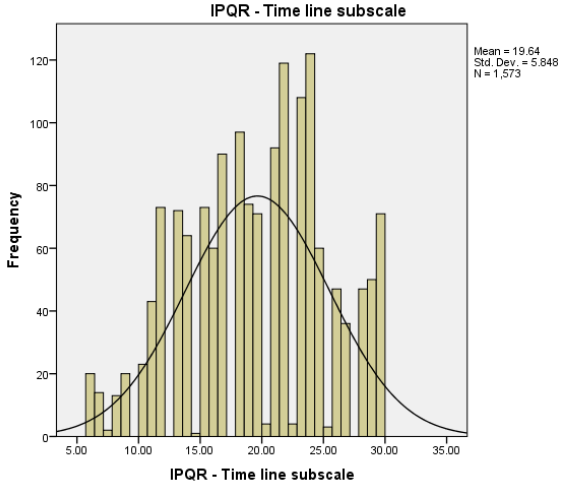

TSK, BL
Mean = 39.67
Std. Dev. = 6.887
N = 1,510
Frequency
TSK, BL

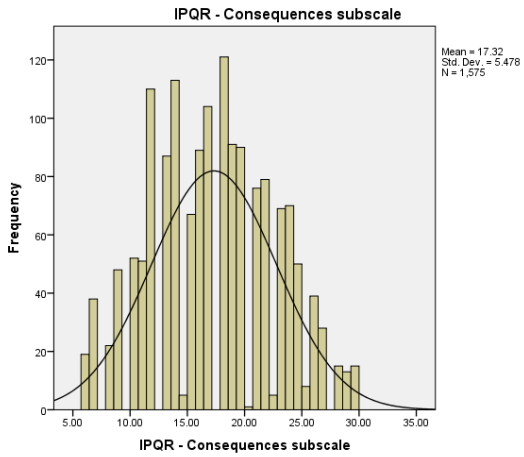

HAD anxiety score, BL
Mean = 8.26
Std. Dev. = 4.554
N = 1,579
Frequency
HAD anxiety score, BL

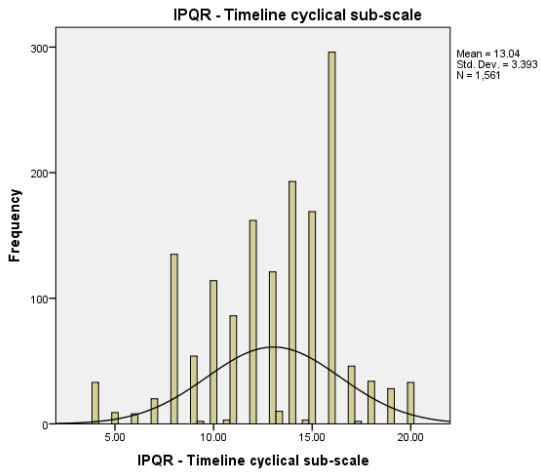

HAD depression score, BL
Mean = 6.53
Std. Dev. = 4.369
N = 1,579
Frequency
HAD depression score, BL

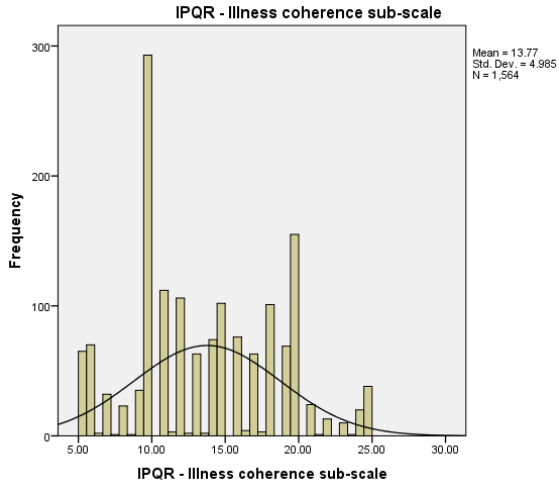

IPQR - symptoms original, BL
Mean = 4.03
Std. Dev. = 2.376
N = 1,591
Frequency
IPQR - symptoms original, BL

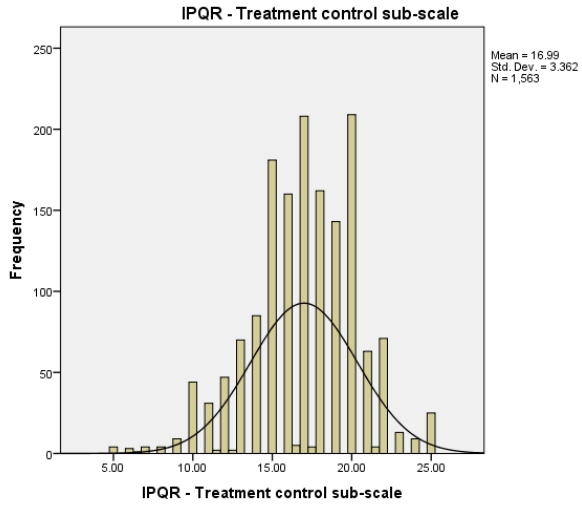

IPQR - Personal control sub-scale
Mean = 20.52
Std. Dev. = 3.788
N = 1,567
Frequency
IPQR - Personal control sub-scale

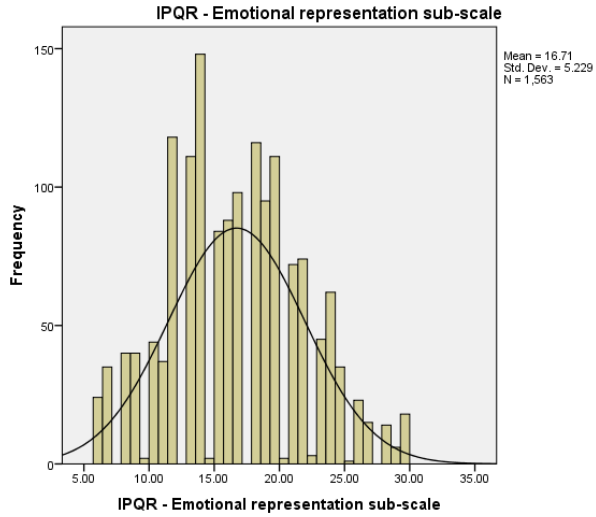

IPQR - Time line subscale
Mean = 19.64
Std. Dev. = 5.848
N = 1,573
Frequency
IPQR - Time line subscale

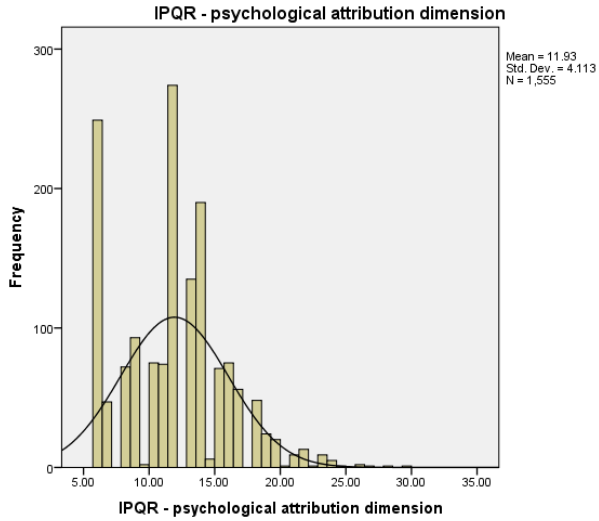

IPQR - Consequences subscale
Mean = 17.32
Std. Dev. = 5.478
N = 1,575
Frequency
IPQR - Consequences subscale

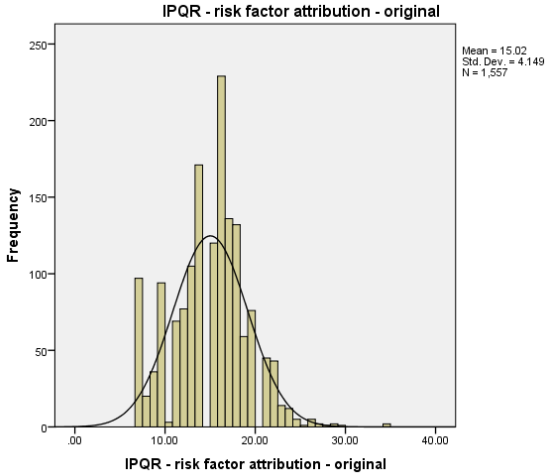

IPQR - Timeline cyclical sub-scale
Mean = 13.04
Std. Dev. = 3.393
N = 1,561
Frequency
IPQR - Timeline cyclical sub-scale

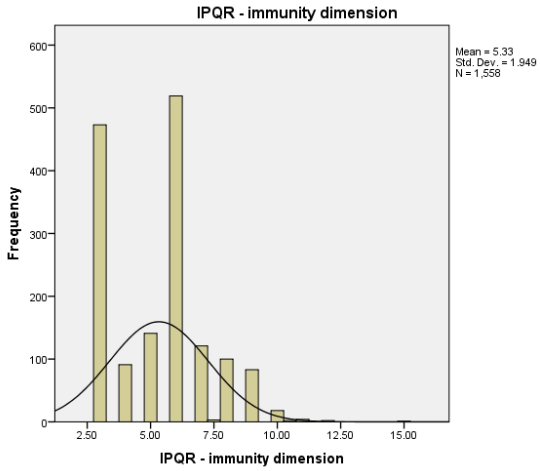

IPQR - Illness coherence sub-scale
Mean = 13.77
Std. Dev. = 4.985
N = 1,564
Frequency
IPQR - Illness coherence sub-scale

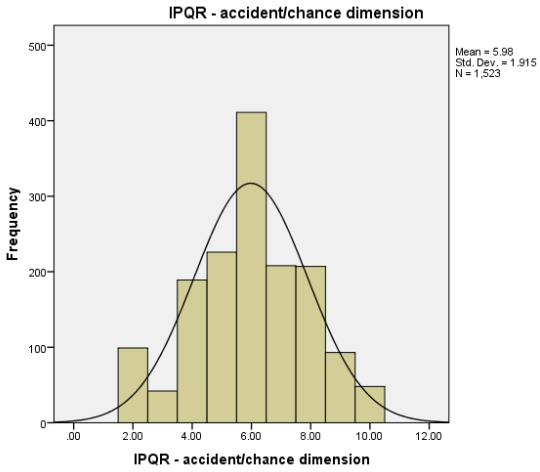

IPQR - Treatment control sub-scale
Mean = 16.99
Std. Dev. = 3.362
N = 1,563
Frequency
IPQR - Treatment control sub-scale

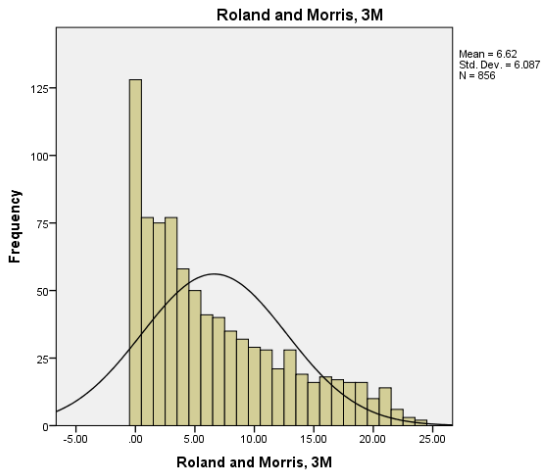

IPQR - Emotional representation sub-scale
Mean = 16.71
Std. Dev. = 5.229
N = 1,563
Frequency
IPQR - Emotional representation sub-scale


IPQR - psychological attribution dimension
Mean = 11.93
Std. Dev. = 4.113
N = 1,555
Frequency
IPQR - psychological attribution dimension

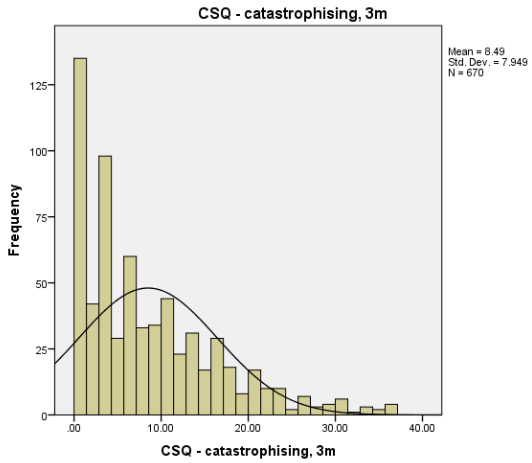

IPQR - risk factor attribution - original
Mean = 15.02
Std. Dev. = 4.149
N = 1,557
Frequency
IPQR - risk factor attribution - original

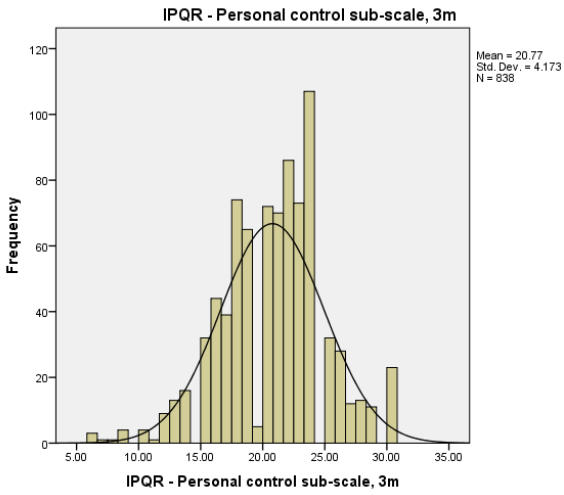

IPQR - immunity dimension
Mean = 5.33
Std. Dev. = 1.949
N = 1,558
Frequency
IPQR - immunity dimension

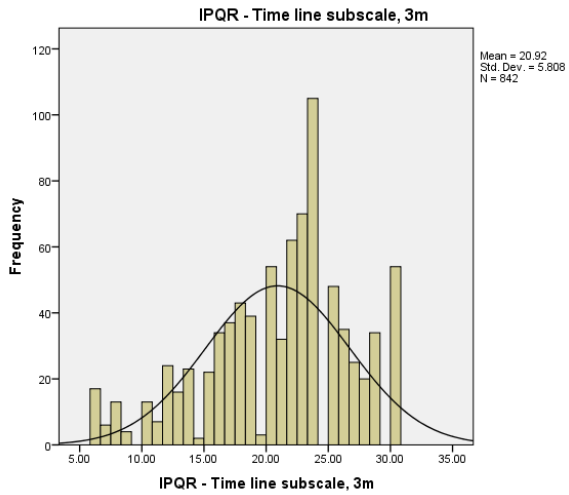

IPQR - accident/chance dimension
Mean = 5.98
Std. Dev. = 1.915
N = 1,523
Frequency
IPQR - accident/chance dimension

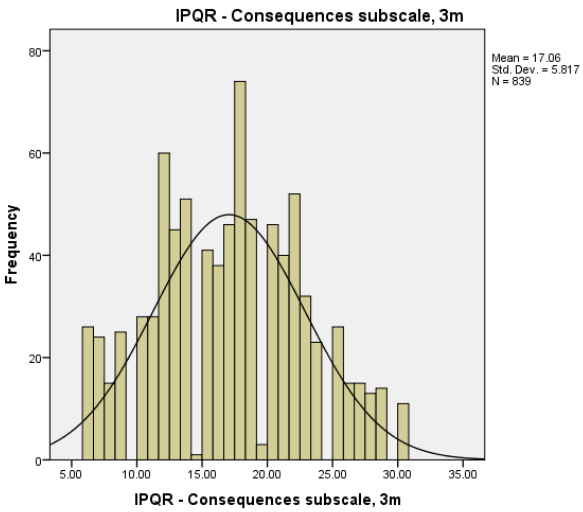

Roland and Morris, 3M
Mean = 6.62
Std. Dev. = 6.087
N = 856
Frequency
Roland and Morris, 3M

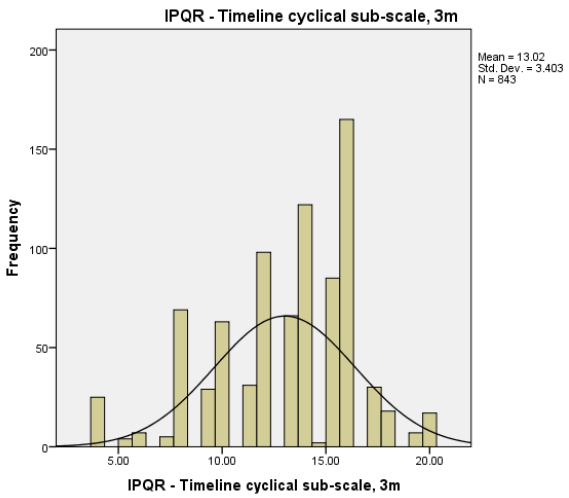

Self efficacy score, 3m
Mean = 39.96
Std. Dev. = 14.11
N = 695
Frequency
Self efficacy score, 3m

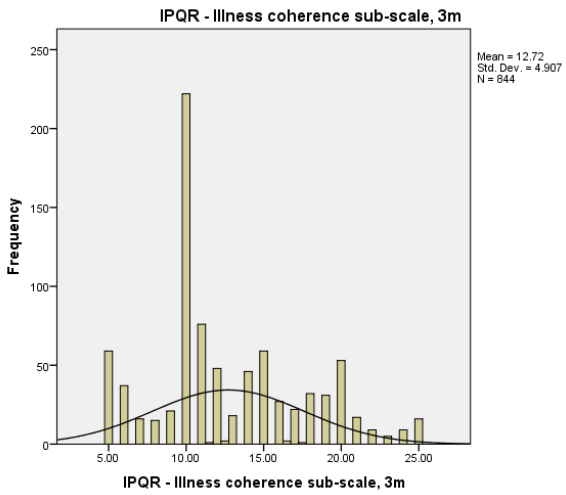

CSQ - catastrophising, 3m
Mean = 8.49
Std. Dev. = 7.949
N = 670
Frequency
CSQ - catastrophising, 3m

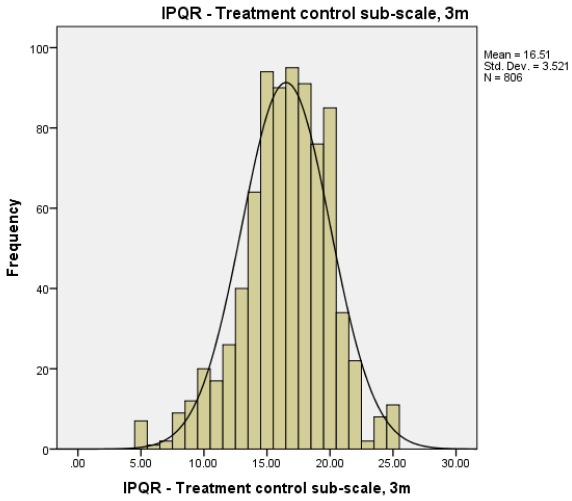

IPQR - Personal control sub-scale, 3m
Mean = 20.77
Std. Dev. = 4.173
N = 838
Frequency
IPQR - Personal control sub-scale, 3m

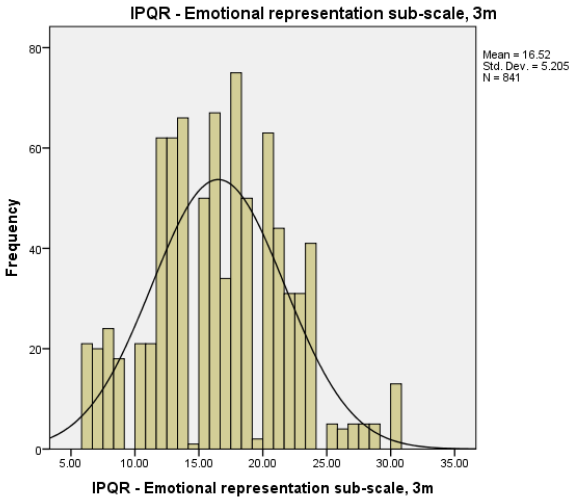

IPQR - Time line subscale, 3m
Mean = 20.92
Std. Dev. = 5.808
N = 842
Frequency
IPQR - Time line subscale, 3m

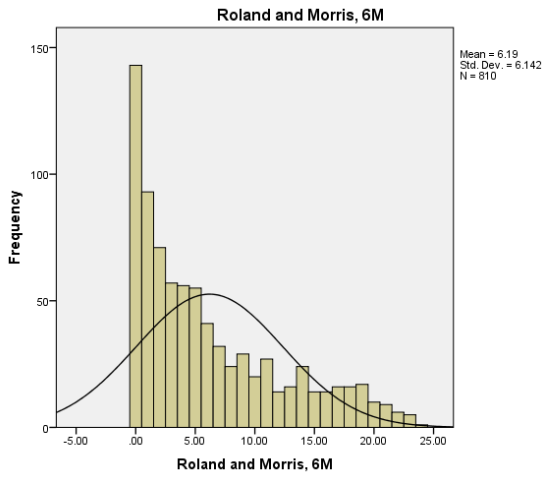

IPQR - Consequences subscale, 3m
Mean = 17.06
Std. Dev. = 5.817
N = 839
Frequency
IPQR - Consequences subscale, 3m

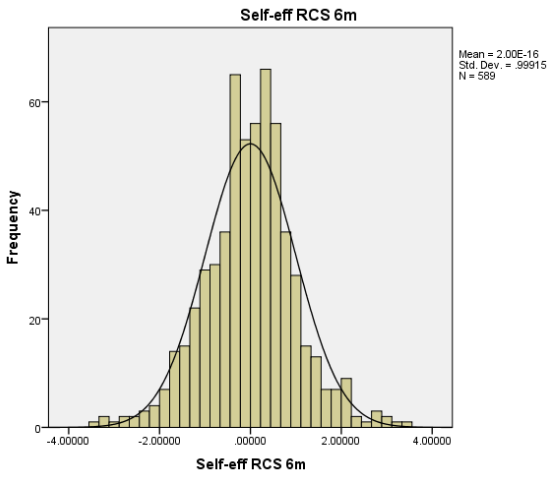

IPQR - Timeline cyclical sub-scale, 3m
Mean = 13.02
Std. Dev. = 3.403
N = 843
Frequency
IPQR - Timeline cyclical sub-scale, 3m

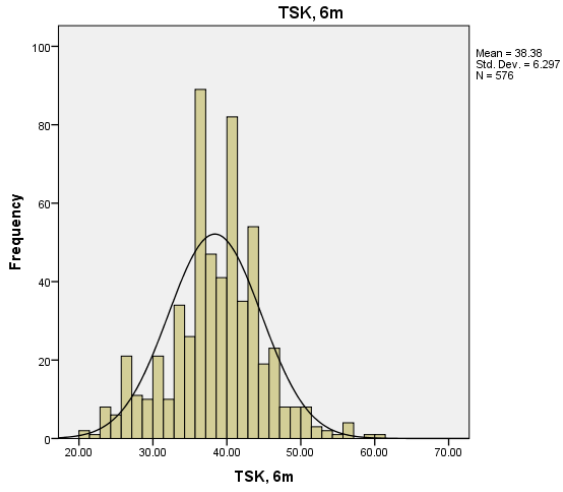

IPQR - Illness coherence sub-scale, 3m
Mean = 12.72
Std. Dev. = 4.907
N = 844
Frequency
IPQR - Illness coherence sub-scale, 3m

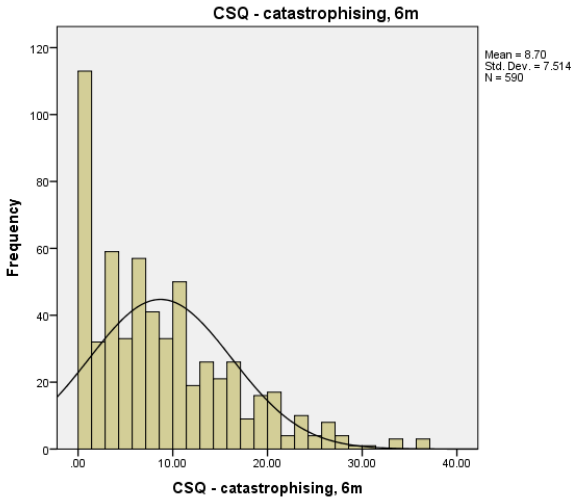

IPQR - Treatment control sub-scale, 3m
Mean = 16.51
Std. Dev. = 3.521
N = 806
Frequency
IPQR - Treatment control sub-scale, 3m

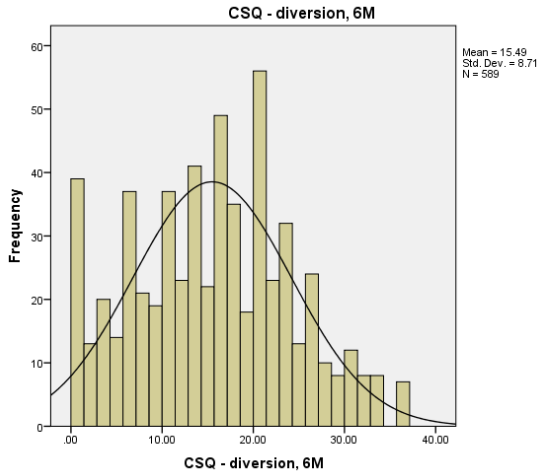

IPQR - Emotional representation sub-scale, 3m
Mean = 16.52
Std. Dev. = 5.205
N = 841
Frequency
IPQR - Emotional representation sub-scale, 3m

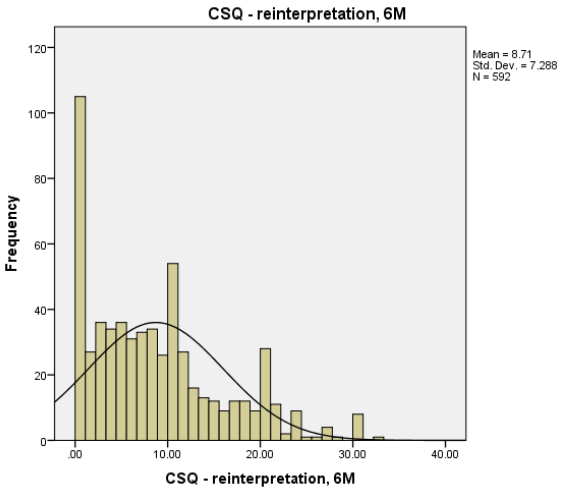

Roland and Morris, 6M
Mean = 6.19
Std. Dev. = 6.142
N = 810
Frequency
Roland and Morris, 6M

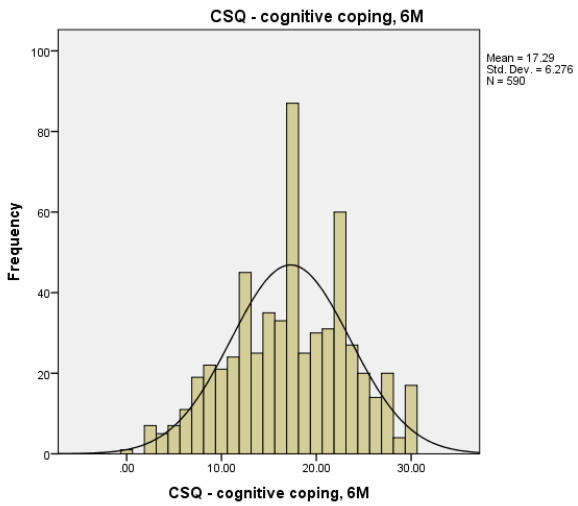

Self-eff RCS 6m
Mean = 2.00E-16
Std. Dev. = .99915
N = 589
Frequency
Self-eff RCS 6m

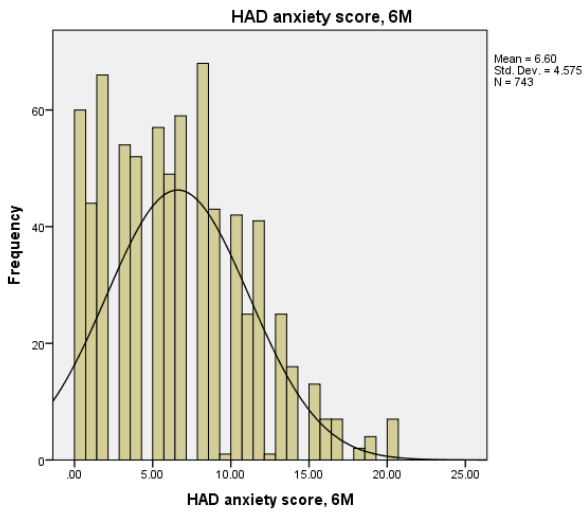

TSK, 6m
Mean = 38.38
Std. Dev. = 6.297
N = 576
Frequency
TSK, 6m

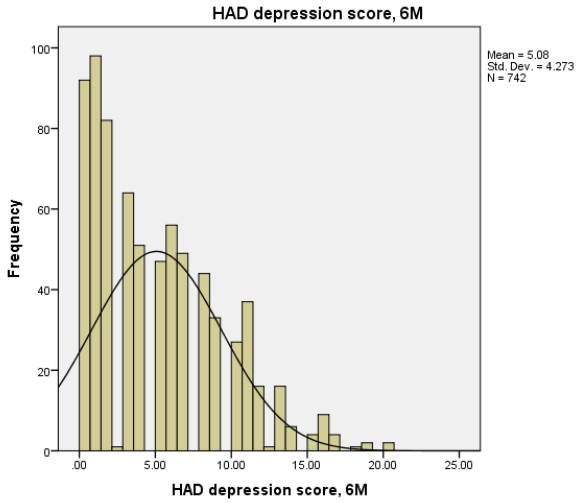

CSQ - catastrophising, 6m
Mean = 8.70
Std. Dev. = 7.514
N = 590
Frequency
CSQ - catastrophising, 6m

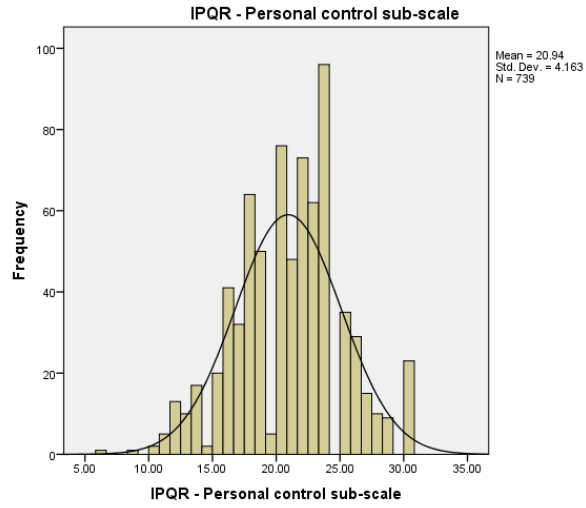

CSQ - diversion, 6M
Mean = 15.49
Std. Dev. = 8.71
N = 589
Frequency
CSQ - diversion, 6M

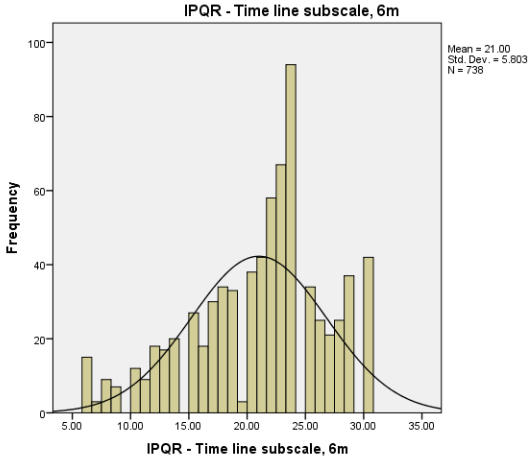

CSQ - reinterpretation, 6M
Mean = 8.71
Std. Dev. = 7.288
N = 592
Frequency
CSQ - reinterpretation, 6M

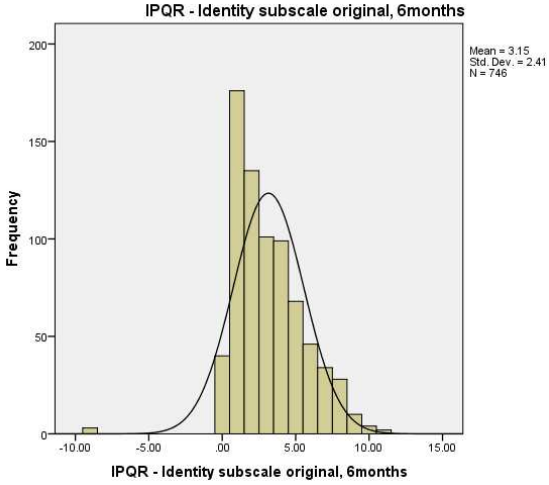

CSQ - cognitive coping, 6M
Mean = 17.29
Std. Dev. = 6.276
N = 590
Frequency
CSQ - cognitive coping, 6M

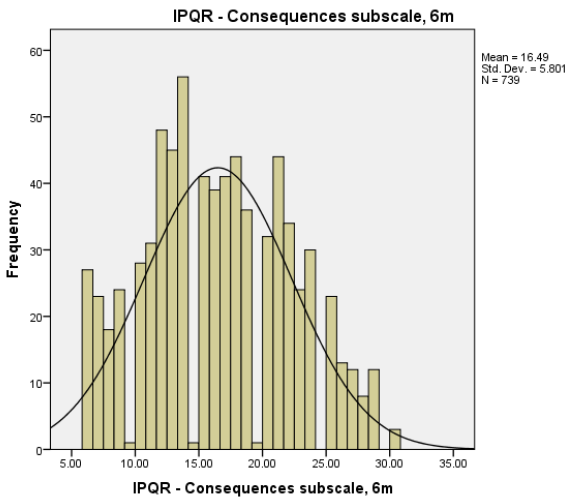

HAD anxiety score, 6M
Mean = 6.60
Std. Dev. = 4.575
N = 743
Frequency
HAD anxiety score, 6M

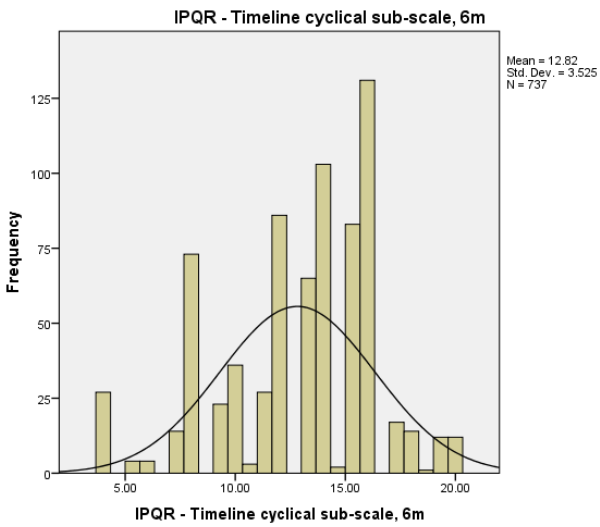

HAD depression score, 6M
Mean = 5.08
Std. Dev. = 4.273
N = 742
Frequency
HAD depression score, 6M

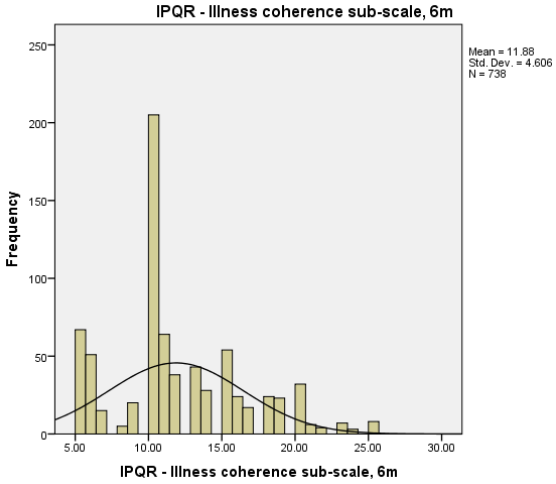

IPQR - Personal control sub-scale
Mean = 20.94
Std. Dev. = 4.163
N = 739
Frequency
IPQR - Personal control sub-scale

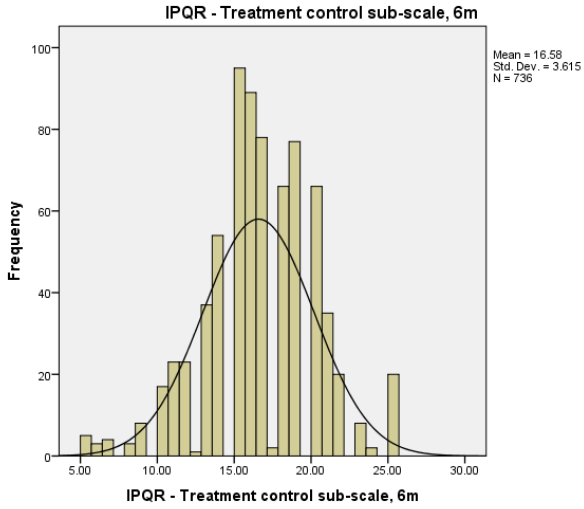

IPQR - Time line subscale, 6m
Mean = 21.00
Std. Dev. = 5.803
N = 738
Frequency
IPQR - Time line subscale, 6m

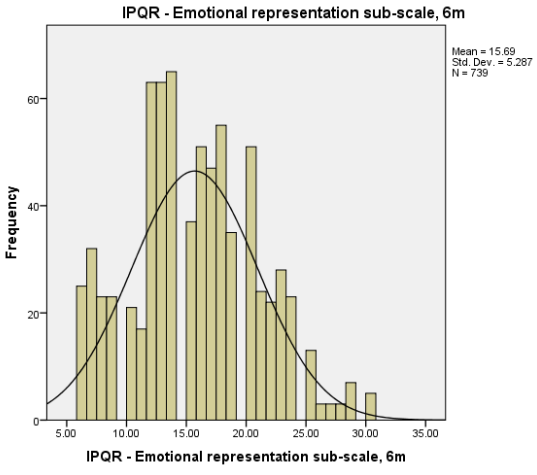

IPQR - Identity subscale original, 6months
Mean = 3.15
Std. Dev. = 2.41
N = 746
Frequency
IPQR - Identity subscale original, 6months

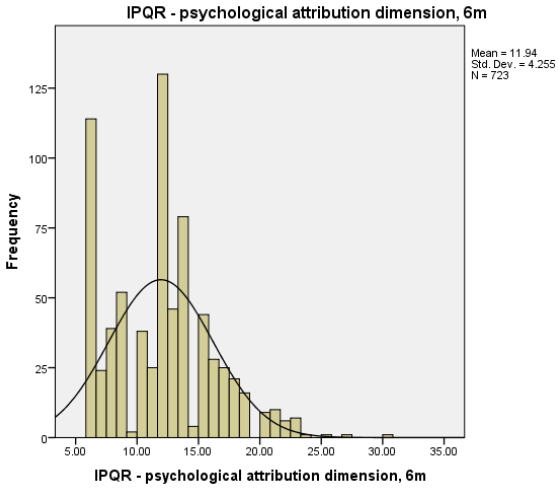

IPQR - Consequences subscale, 6m
Mean = 16.49
Std. Dev. = 5.801
N = 739
Frequency
IPQR - Consequences subscale, 6m

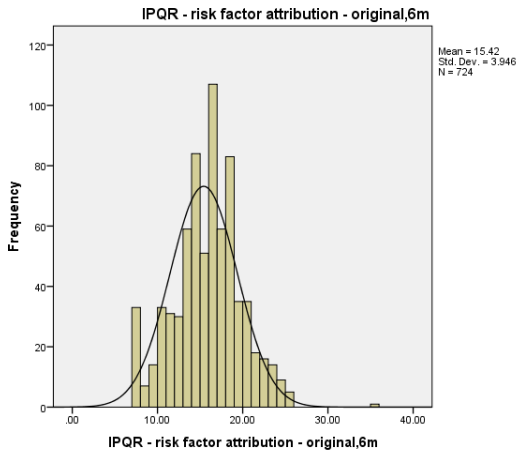

IPQR - Timeline cyclical sub-scale, 6m
Mean = 12.82
Std. Dev. = 3.525
N = 737
Frequency
IPQR - Timeline cyclical sub-scale, 6m

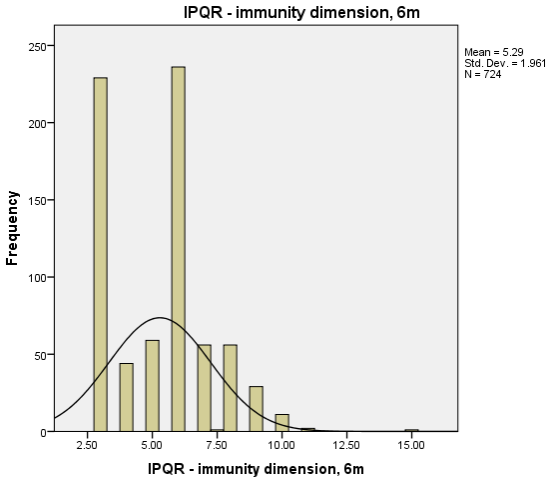

IPQR - Illness coherence sub-scale, 6m
Mean = 11.88
Std. Dev. = 4.606
N = 738
Frequency
IPQR - Illness coherence sub-scale, 6m

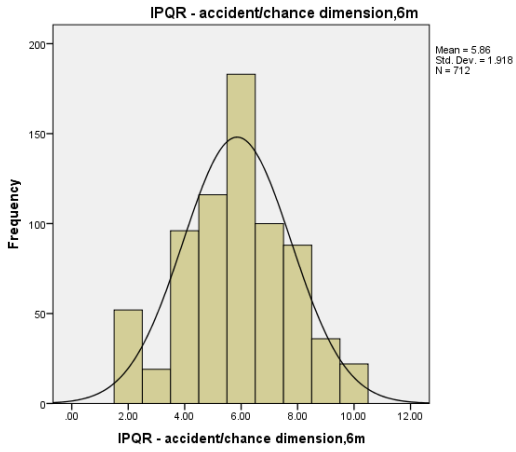

IPQR - Treatment control sub-scale, 6m
Mean = 16.58
Std. Dev. = 3.615
N = 736
Frequency
IPQR - Treatment control sub-scale, 6m

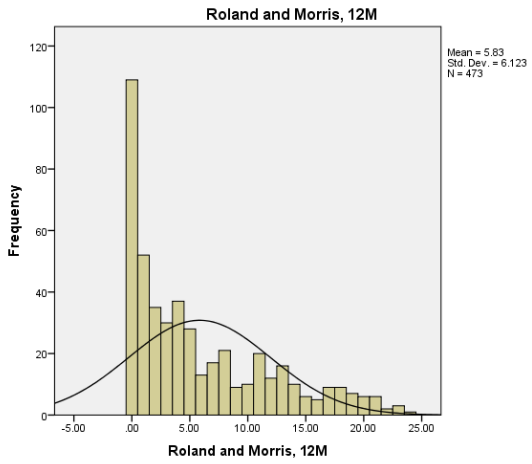

IPQR - Emotional representation sub-scale, 6m
Mean = 15.69
Std. Dev. = 5.287
N = 739
Frequency
IPQR - Emotional representation sub-scale, 6m

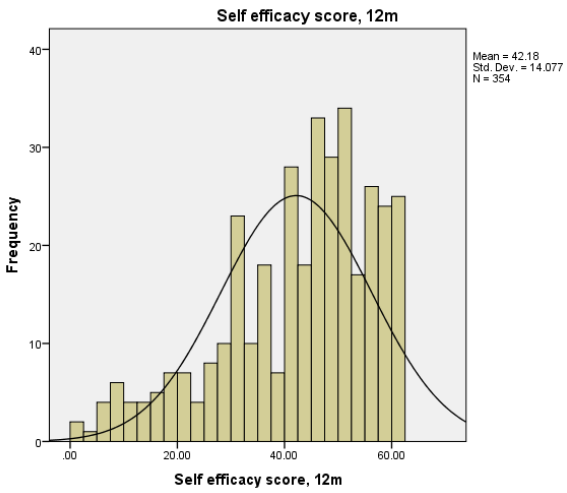

IPQR - psychological attribution dimension, 6m
Mean = 11.94
Std. Dev. = 4.255
N = 723
Frequency
IPQR - psychological attribution dimension, 6m

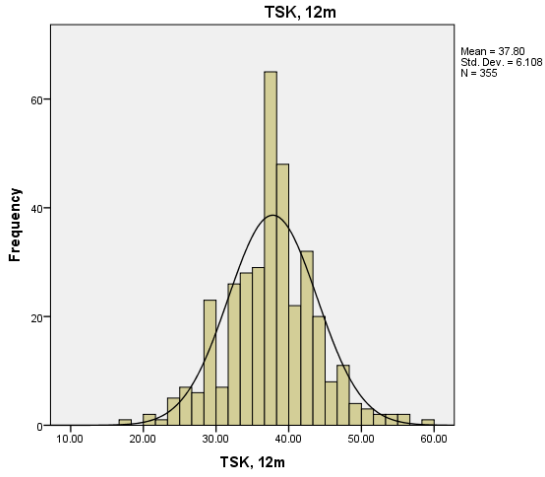

IPQR - risk factor attribution - original,6m
Mean = 15.42
Std. Dev. = 3.946
N = 724
Frequency
IPQR - risk factor attribution - original,6m

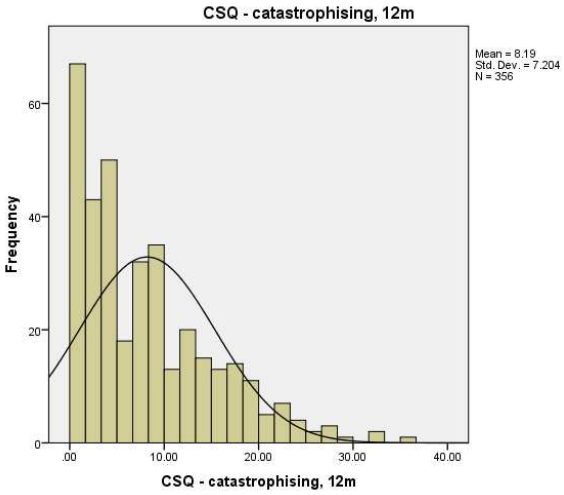

IPQR - immunity dimension, 6m
Mean = 5.29
Std. Dev. = 1.961
N = 724
Frequency
IPQR - immunity dimension, 6m

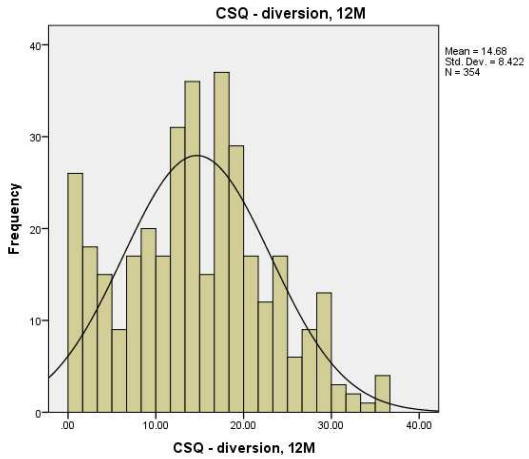

IPQR - accident/chance dimension,6m
Mean = 5.86
Std. Dev. = 1.918
N = 712
Frequency
IPQR - accident/chance dimension,6m

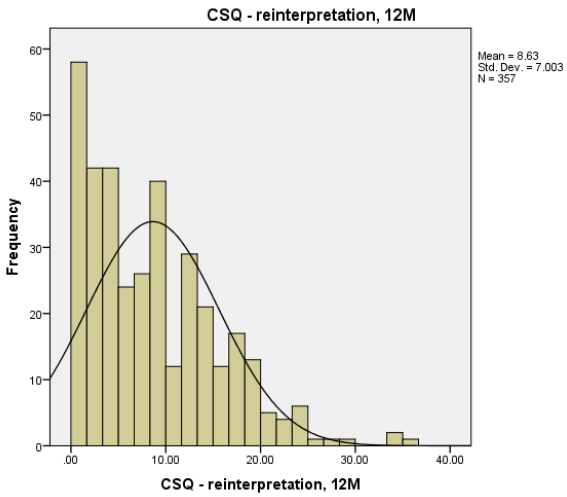

Roland and Morris, 12M
Mean = 5.83
Std. Dev. = 6.123
N = 473
Frequency
Roland and Morris, 12M

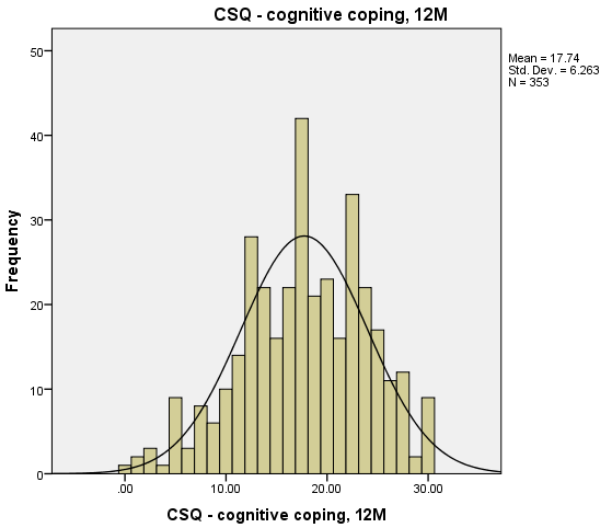

Self efficacy score, 12m
Mean = 42.18
Std. Dev. = 14.077
N = 354
Frequency
Self efficacy score, 12m

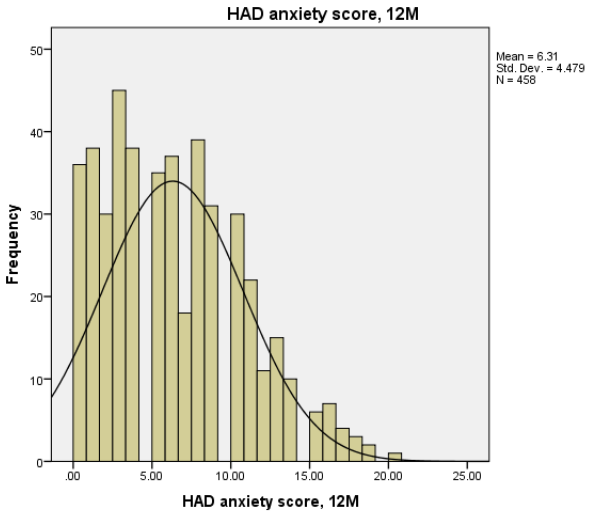

TSK, 12m
Mean = 37.80
Std. Dev. = 6.108
N = 355
Frequency
TSK, 12m

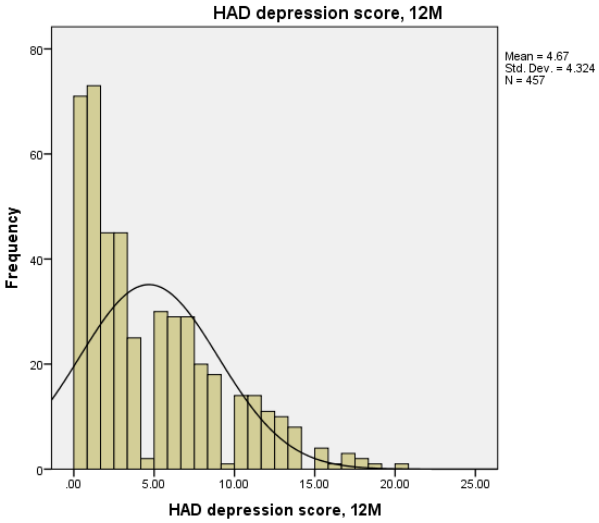

CSQ - catastrophising, 12m
Mean = 8.19
Std. Dev. = 7.204
N = 356
Frequency
CSQ - catastrophising, 12m

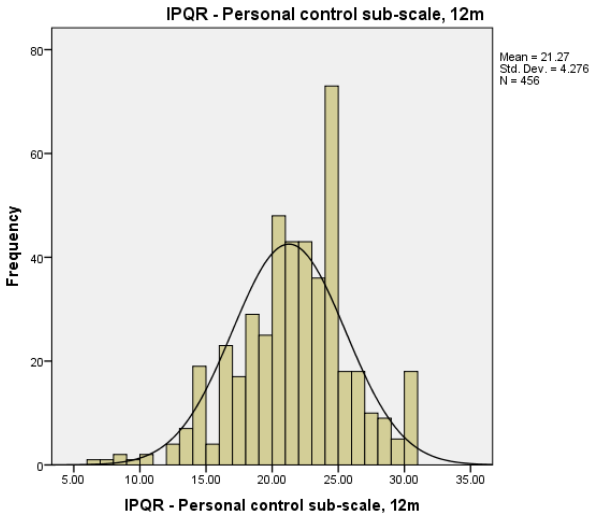

CSQ - diversion, 12M
Mean = 14.68
Std. Dev. = 8.422
N = 354
Frequency
CSQ - diversion, 12M

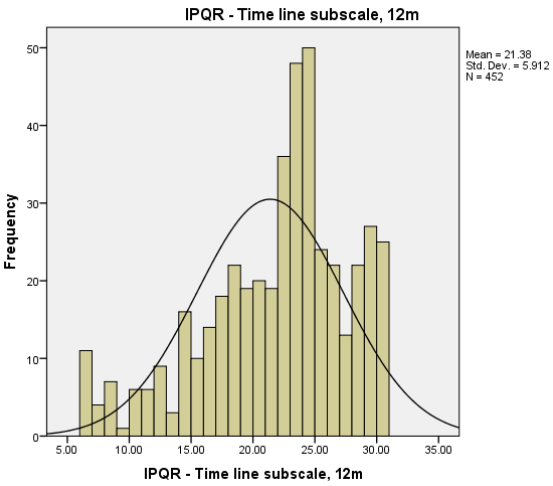

CSQ - reinterpretation, 12M
Mean = 8.63
Std. Dev. = 7.003
N = 357
Frequency
CSQ - reinterpretation, 12M

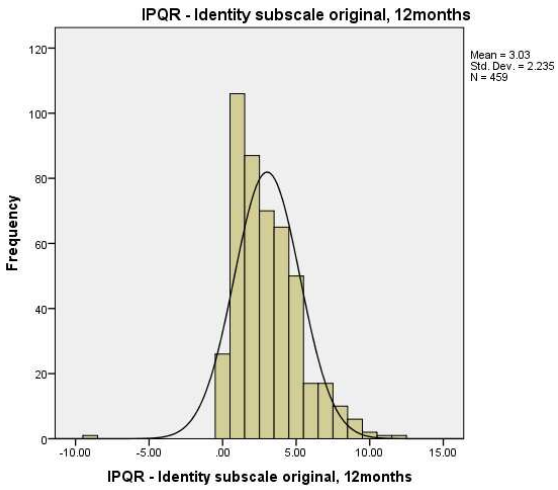

CSQ - cognitive coping, 12M
Mean = 17.74
Std. Dev. = 6.263
N = 353
Frequency
CSQ - cognitive coping, 12M

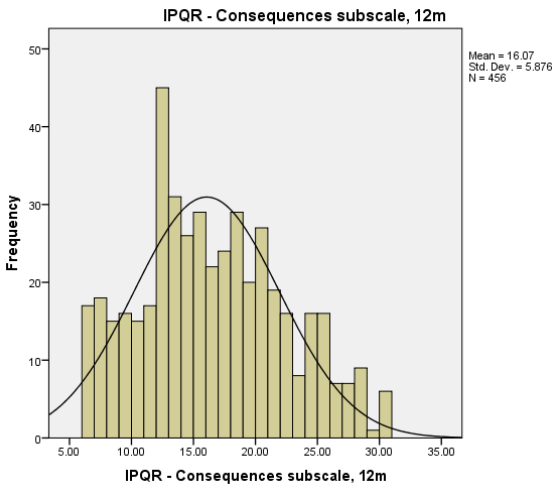

HAD anxiety score, 12M
Mean = 6.31
Std. Dev. = 4.479
N = 458
Frequency
HAD anxiety score, 12M

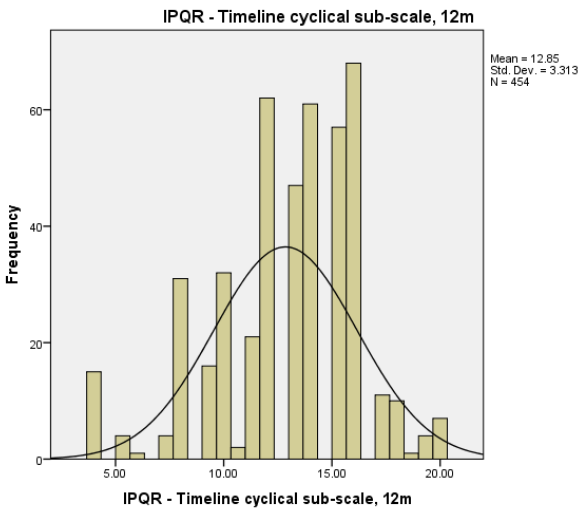

HAD depression score, 12M
Mean = 4.67
Std. Dev. = 4.324
N = 457
Frequency
HAD depression score, 12M

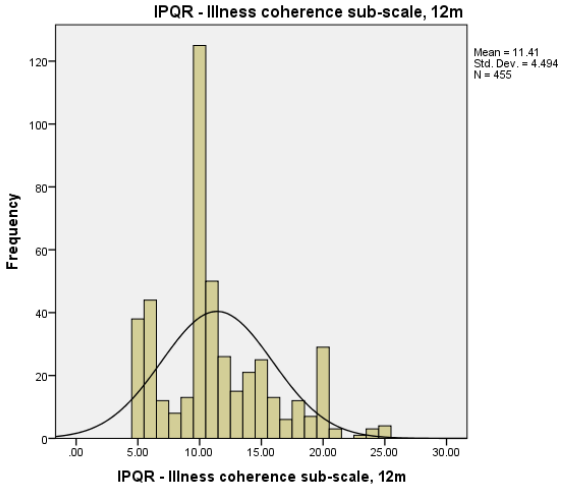

IPQR - Personal control sub-scale, 12m
Mean = 21.27
Std. Dev. = 4.276
N = 456
Frequency
IPQR - Personal control sub-scale, 12m

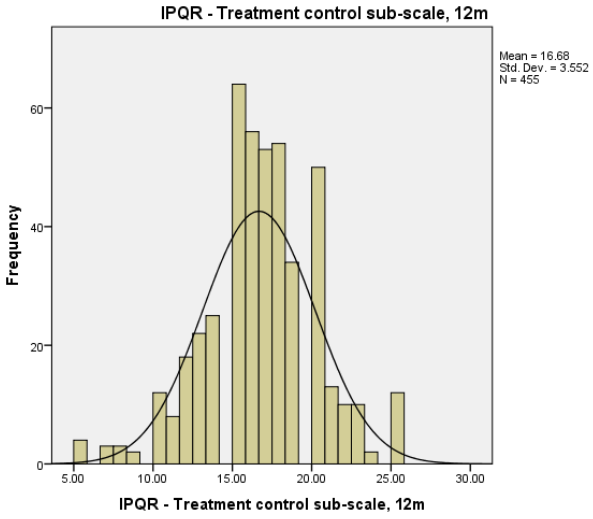

IPQR - Time line subscale, 12m
Mean = 21.38
Std. Dev. = 5.912
N = 452
Frequency
IPQR - Time line subscale, 12m

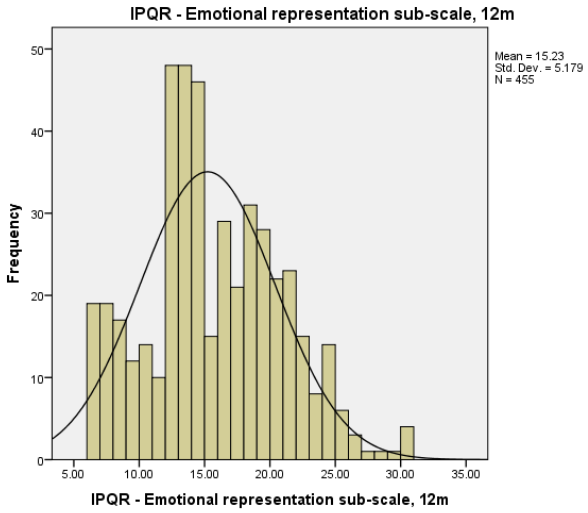

IPQR - Identity subscale original, 12months
Mean = 3.03
Std. Dev. = 2.235
N = 459
Frequency
IPQR - Identity subscale original, 12months

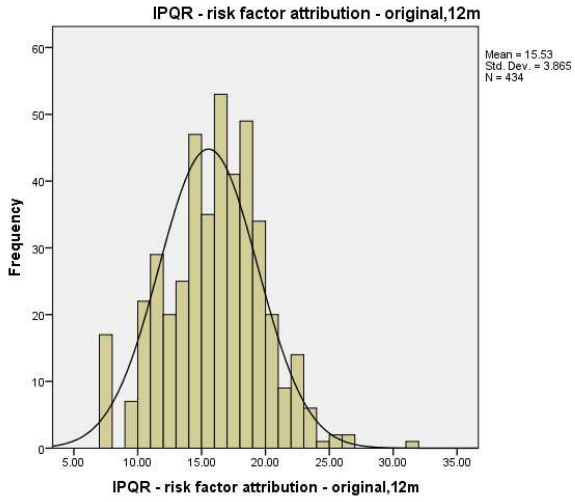

IPQR - Consequences subscale, 12m
Mean = 16.07
Std. Dev. = 5.876
N = 456
Frequency
IPQR - Consequences subscale, 12m

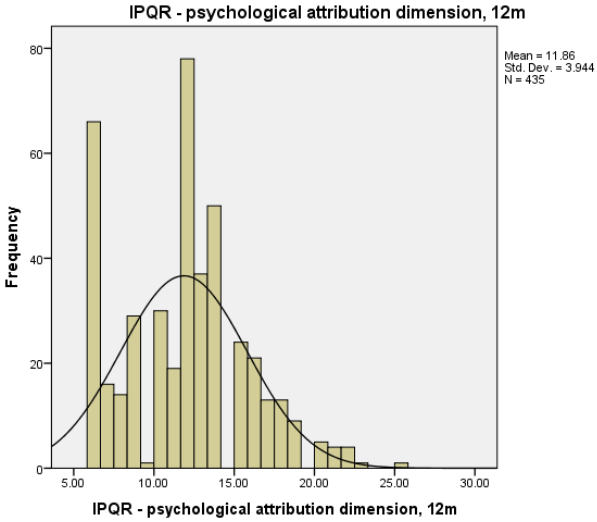

IPQR - Timeline cyclical sub-scale, 12m
Mean = 12.85
Std. Dev. = 3.313
N = 454
Frequency
IPQR - Timeline cyclical sub-scale, 12m

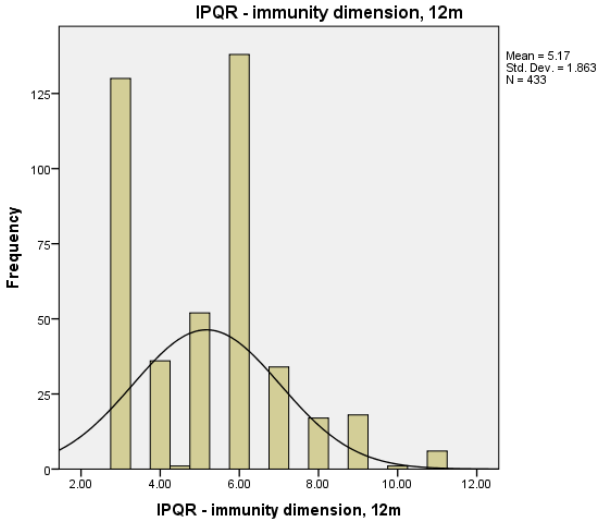

IPQR - Illness coherence sub-scale, 12m
Mean = 11.41
Std. Dev. = 4.494
N = 455
Frequency
IPQR - Illness coherence sub-scale, 12m

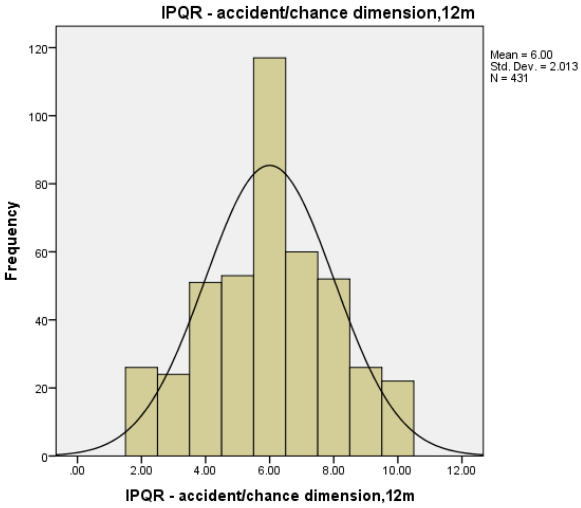

IPQR - Treatment control sub-scale, 12m
Mean = 16.68
Std. Dev. = 3.552
N = 455
Frequency
IPQR - Treatment control sub-scale, 12m


IPQR - Emotional representation sub-scale, 12m
Mean = 15.23
Std. Dev. = 5.179
N = 455
Frequency
IPQR - Emotional representation sub-scale, 12m


IPQR - risk factor attribution - original,12m
Mean = 15.53
Std. Dev. = 3.865
N = 434
Frequency
IPQR - risk factor attribution - original,12m

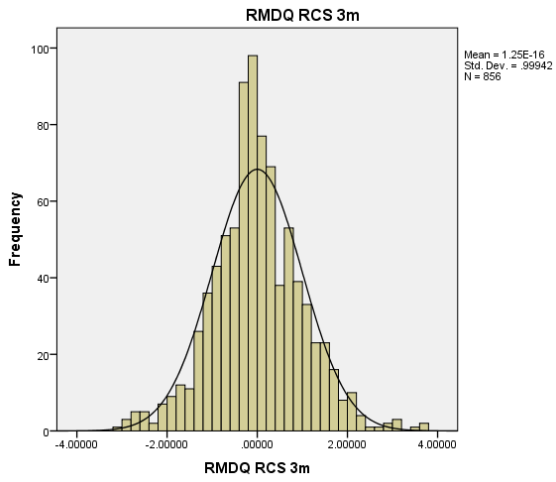

IPQR - psychological attribution dimension, 12m
Mean = 11.86
Std. Dev. = 3.944
N = 435
Frequency
IPQR - psychological attribution dimension, 12m

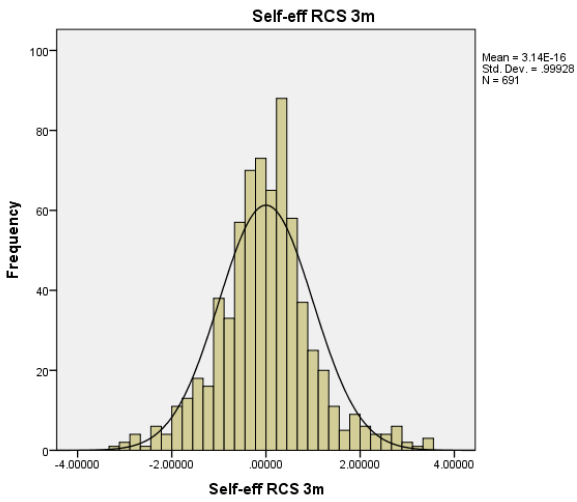

IPQR - immunity dimension, 12m
Mean = 5.17
Std. Dev. = 1.863
N = 433
Frequency
IPQR - immunity dimension, 12m

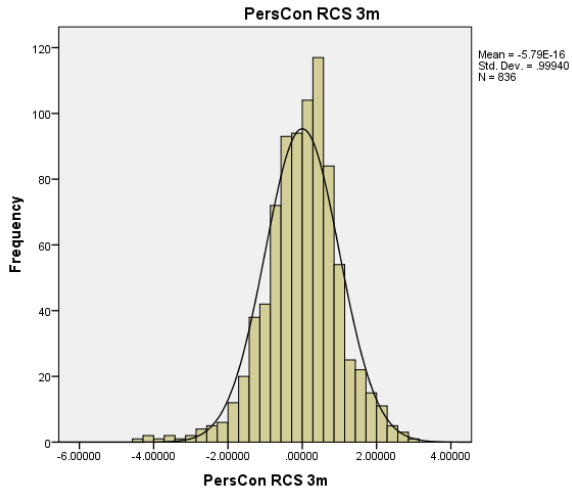

IPQR - accident/chance dimension,12m
Mean = 6.00
Std. Dev. = 2.013
N = 431
Frequency
IPQR - accident/chance dimension,12m

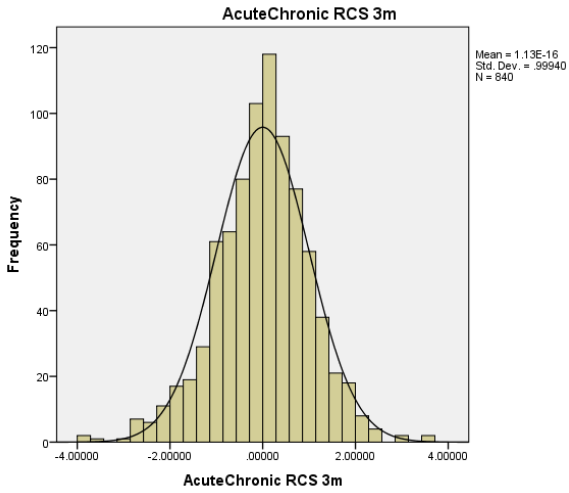
Residualised change scores

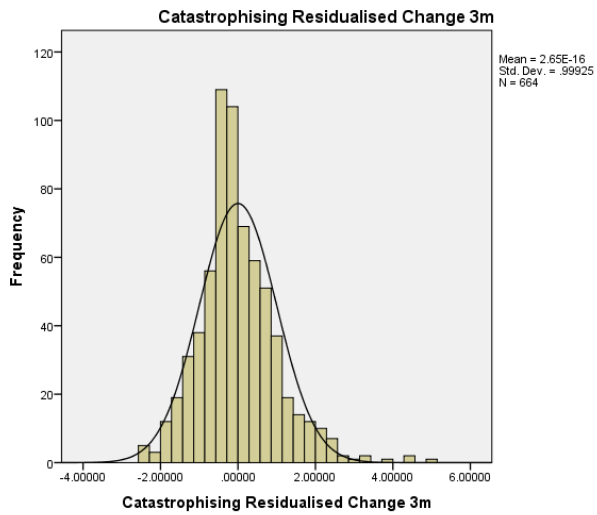
		Baseline to three-month residualised change		Baseline to six-month residualised change		Baseline to 12-month residualised change	
Measure		**Skewness (SE)**	**Kurtosis (SE)**	**Skewness (SE)**	**Kurtosis (SE)**	**Skewness (SE)**	**Kurtosis (SE)**
RMDQ		0.14 (0.08)	0.88 (0.17)	0.03 (0.09)	0.60 (0.17)	-0.06 (0.11)	0.40 (0.22)
Self-efficacy		0.22 (0.09)	1.25 (0.19)	0.02 (0.10)	0.90 (0.20)	0.01 (0.13)	0.60 (0.26)
Coping	Catastrophising	0.85 (0.10)	2.37 (0.19)	0.54 (0.10)	2.33 (0.20)	0.77 (0.13)	1.08 (0.26)
	Diversion	***	***	0.06 (0.10)	0.35 (0.20)	0.20 (0.13)	0.15 (0.26)
	Re-interpretation	***	***	0.61 (0.10)	0.79 (0.20)	0.48 (0.13)	0.32 (0.26)
	Cognitive coping	***	***	-0.16 (0.10)	0.79 (0.20)	-0.12 (0.13)	0.06 (0.26)
Fear-avoidance beliefs		***	***	-0.47 (0.10)	1.01 (0.21)	-0.28 (0.13)	0.80 (0.26)
Anxiety		***	***	0.10 (0.09)	0.94 (0.18)	0.20 (0.11)	0.51 (0.23)
Depression		***	***	0.00 (0.09)	0.70 (0.18)	0.48 (0.11)	1.62 (0.23)
Illness perceptions	Personal control	-0.48 (0.09)	1.58 (0.17)	-0.06 (0.09)	0.98 (0.18)	-0.34 (0.12)	1.55 (0.23)
	Acute-chronic timeline	-0.24 (0.08)	0.84 (0.17)	-0.32 (0.09)	0.32 (0.18)	-0.38 (0.12)	0.58 (0.23)
	Identity	***	***	-0.74 (0.09)	5.25 (0.18)	-0.27 (0.11)	4.35 (0.23)
	Consequences	-0.08 (0.09)	0.99 (0.17)	-0.05 (0.09)	0.50 (0.18)	0.08 (0.11)	0.56 (0.23)
	Timeline cyclical	-0.53 (0.09)	0.98 (0.17)	-0.51 (0.09)	0.50 (0.18)	-0.38 (0.12)	0.18 (0.23)
	Illness coherence	0.15 (0.08)	0.55 (0.17)	0.28 (0.09)	0.29 (0.18)	0.30 (0.12)	0.44 (0.23)
	Treatment control	-0.67 (0.09)	3.00 (0.17)	-0.33 (0.09)	1.05 (0.18)	-0.53 (0.12)	2.39 (0.23)
	Emotional representation	-0.13 (0.09)	0.57 (0.17)	-0.07 (0.09)	1.58 (0.18)	0.10 (0.12)	0.43 (0.23)
	Psychological attribution	***	***	0.39 (0.09)	1.48 (0.18)	0.25 (0.12)	-0.03 (0.24)
	Risk factors	***	***	0.14 (0.09)	2.43 (0.18)	0.20 (0.12)	0.83 (0.24)
	Immunity	***	***	0.42 (0.09)	1.23 (0.18)	0.61 (0.12)	1.15 (0.24)
	Accident/Chance	***	***	-0.20 (0.09)	0.43 (0.19)	-0.24 (0.12)	0.14 (0.24)

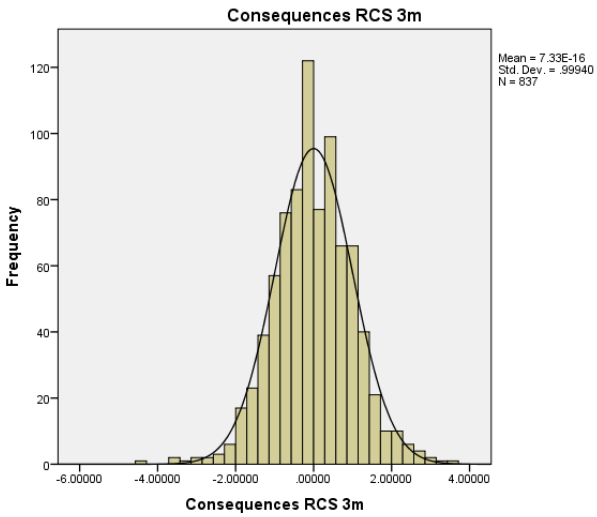

RMDQ RCS 3m
Mean = 1.25E-16
Std. Dev. = .99942
N = 856
Frequency
RMDQ RCS 3m

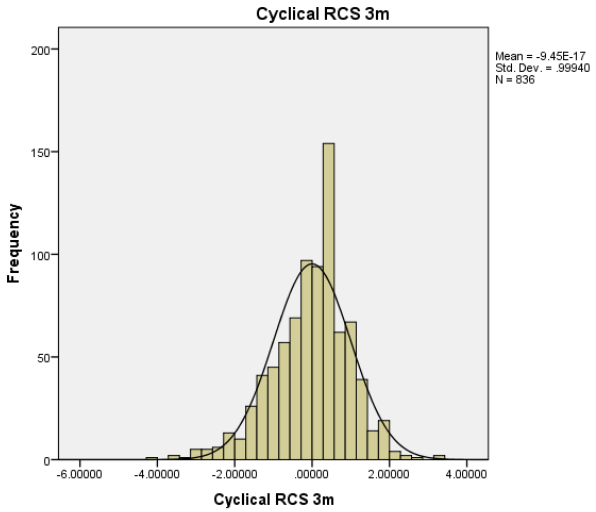

Self-eff RCS 3m
Mean = 3.14E-16
Std. Dev. = .99928
N = 691
Frequency
Self-eff RCS 3m

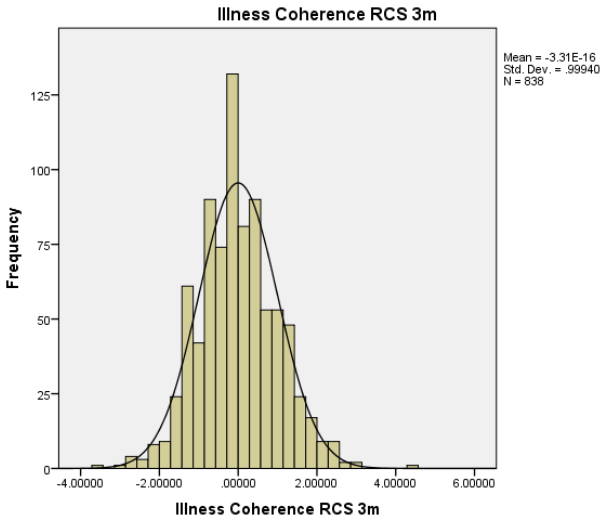

PersCon RCS 3m
Mean = -5.79E-16
Std. Dev. = .99940
N = 836
Frequency
PersCon RCS 3m

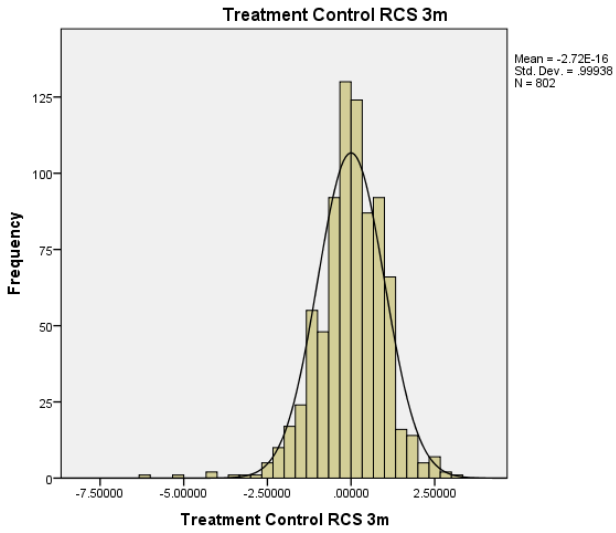

AcuteChronic RCS 3m
Mean = 1.13E-16
Std. Dev. = .99940
N = 840
Frequency
AcuteChronic RCS 3m

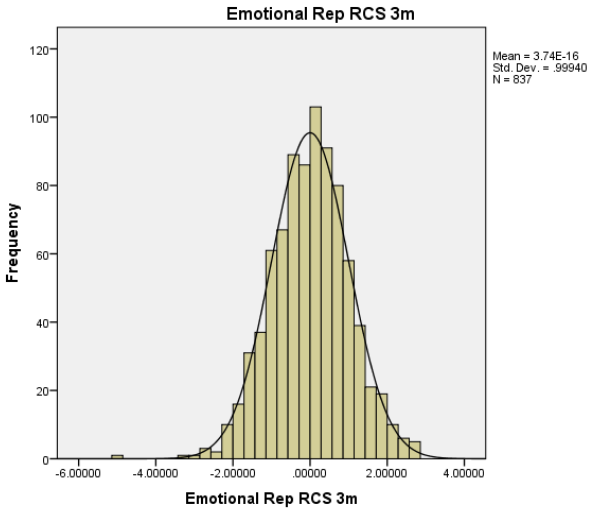

Catastrophising Residualised Change 3m
Mean = 2.65E-16
Std. Dev. = .99925
N = 664
Frequency
Catastrophising Residualised Change 3m

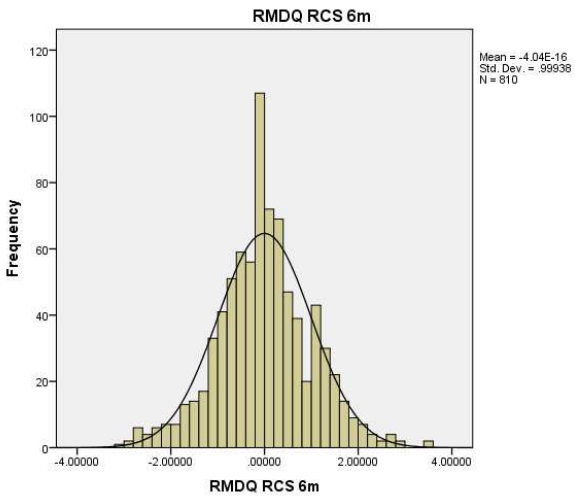

Consequences RCS 3m
Mean = 7.33E-16
Std. Dev. = .99940
N = 837
Frequency
Consequences RCS 3m

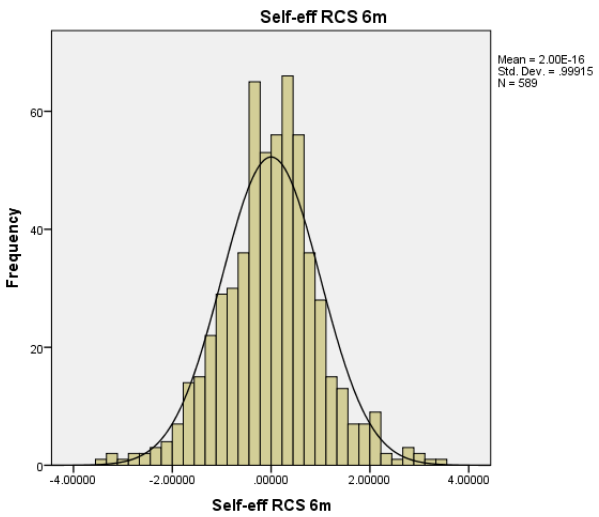

Cyclical RCS 3m
Mean = -9.45E-17
Std. Dev. = .99940
N = 836
Frequency
Cyclical RCS 3m

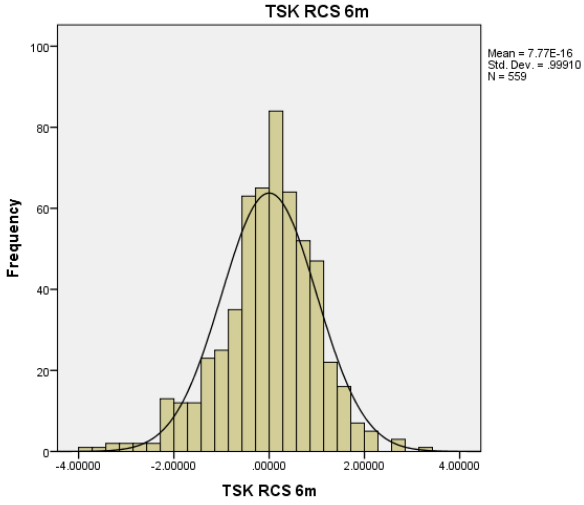

Illness Coherence RCS 3m
Mean = -3.31E-16
Std. Dev. = .99940
N = 838
Frequency
Illness Coherence RCS 3m

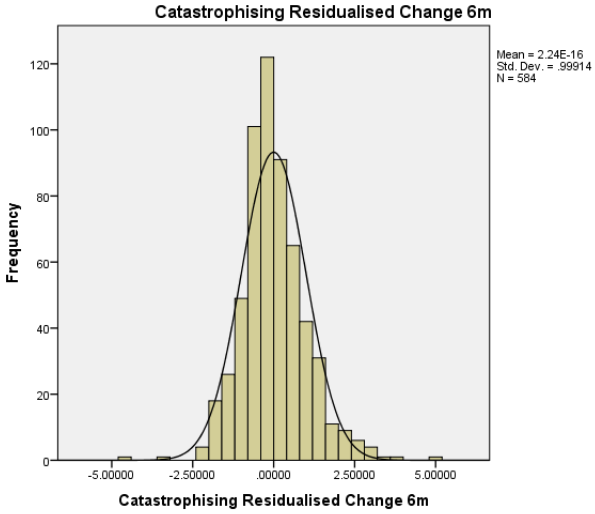

Treatment Control RCS 3m
Mean = -2.72E-16
Std. Dev. = .99938
N = 802
Frequency
Treatment Control RCS 3m

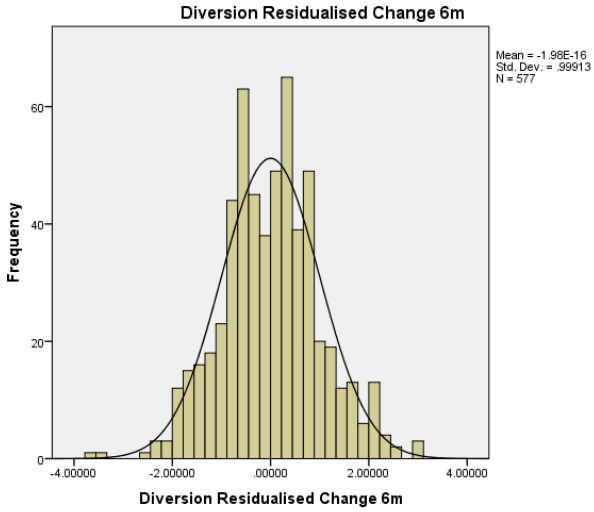

Emotional Rep RCS 3m
Mean = 3.74E-16
Std. Dev. = .99940
N = 837
Frequency
Emotional Rep RCS 3m

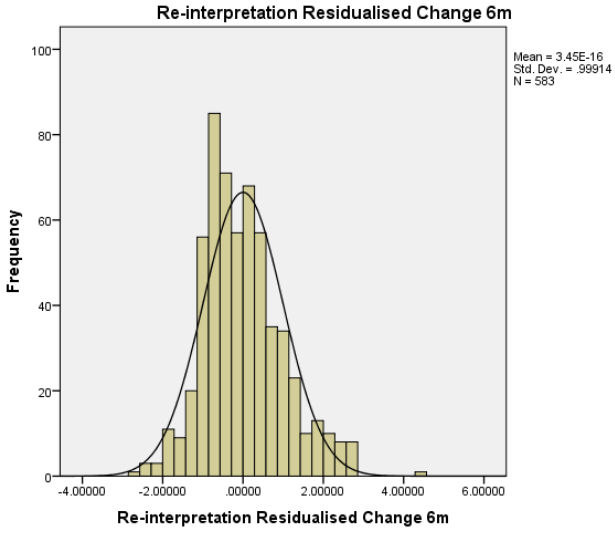

RMDQ RCS 6m
Mean = -4.04E-16
Std. Dev. = .99938
N = 810
Frequency
RMDQ RCS 6m

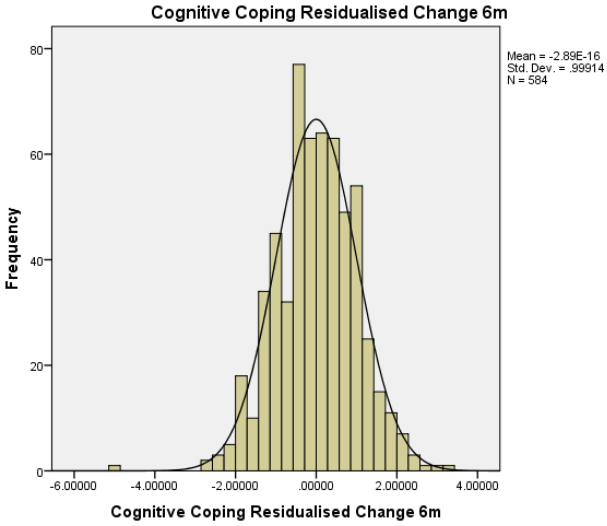

Self-eff RCS 6m
Mean = 2.00E-16
Std. Dev. = .99915
N = 589
Frequency
Self-eff RCS 6m

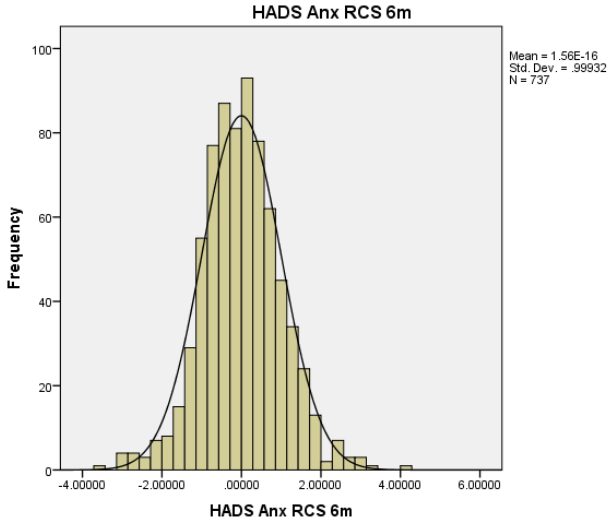

TSK RCS 6m
Mean = 7.77E-16
Std. Dev. = .99910
N = 559
Frequency
TSK RCS 6m

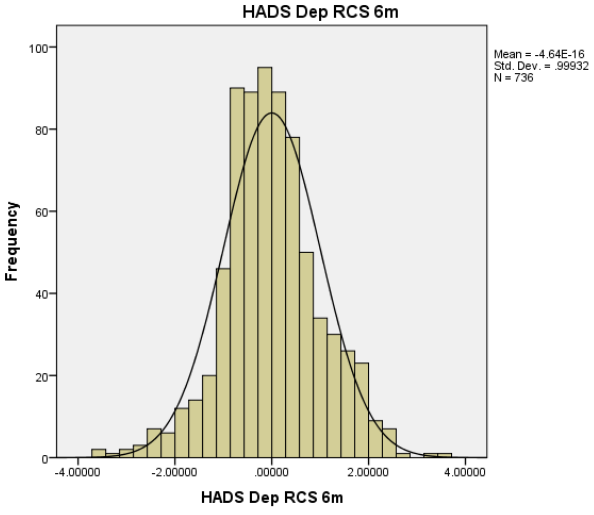

Catastrophising Residualised Change 6m
Mean = 2.24E-16
Std. Dev. = .99914
N = 584
Frequency
Catastrophising Residualised Change 6m

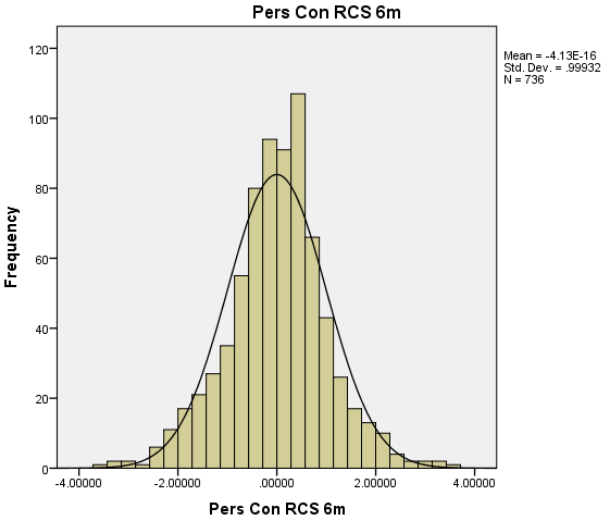

Diversion Residualised Change 6m
Mean = -1.98E-16
Std. Dev. = .99913
N = 577
Frequency
Diversion Residualised Change 6m

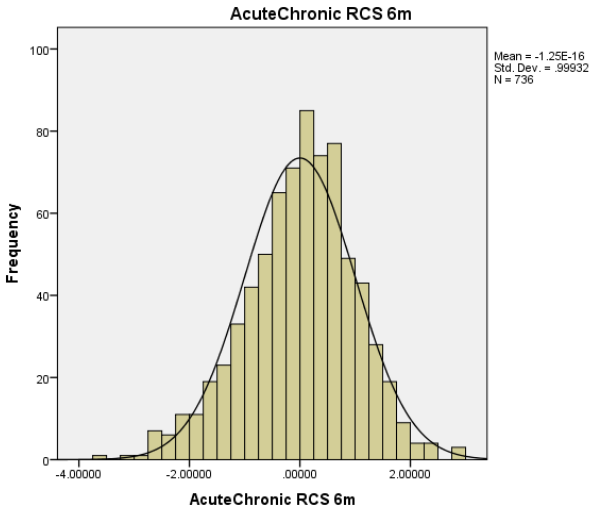

Re-interpretation Residualised Change 6m
Mean = 3.45E-16
Std. Dev. = .99914
N = 583
Frequency
Re-interpretation Residualised Change 6m

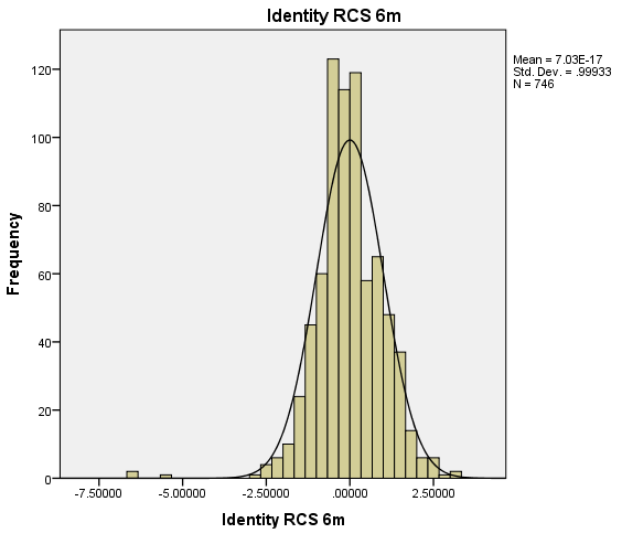

Cognitive Coping Residualised Change 6m
Mean = -2.89E-16
Std. Dev. = .99914
N = 584
Frequency
Cognitive Coping Residualised Change 6m

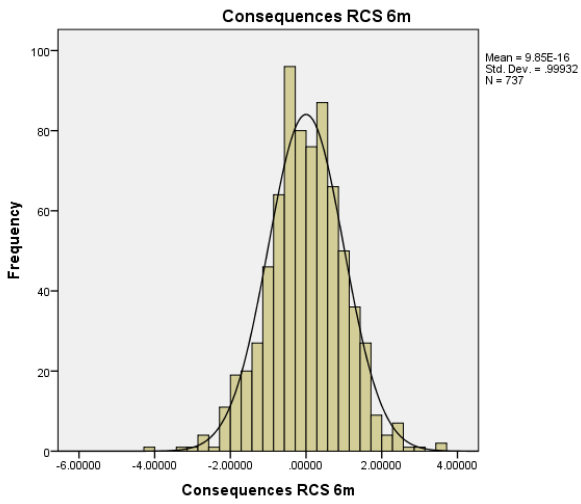

HADS Anx RCS 6m
Mean = 1.56E-16
Std. Dev. = .99932
N = 737
Frequency
HADS Anx RCS 6m

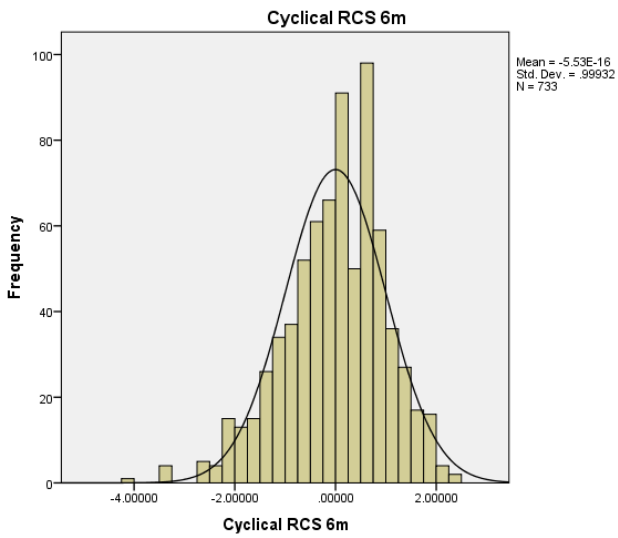

HADS Dep RCS 6m
Mean = -4.64E-16
Std. Dev. = .99932
N = 736
Frequency
HADS Dep RCS 6m

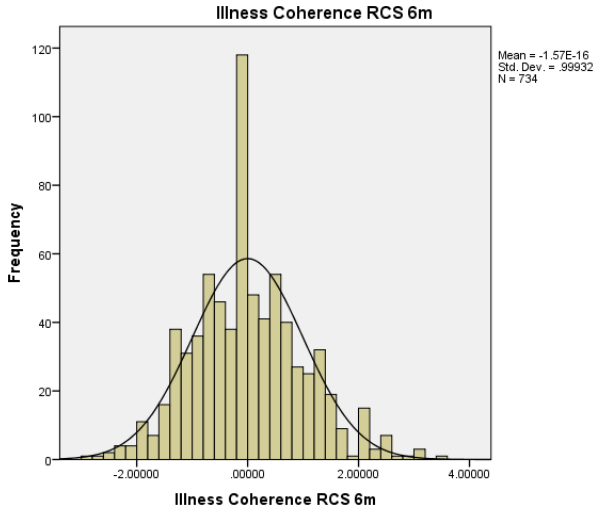

Pers Con RCS 6m
Mean = -4.13E-16
Std. Dev. = .99932
N = 736
Frequency
Pers Con RCS 6m

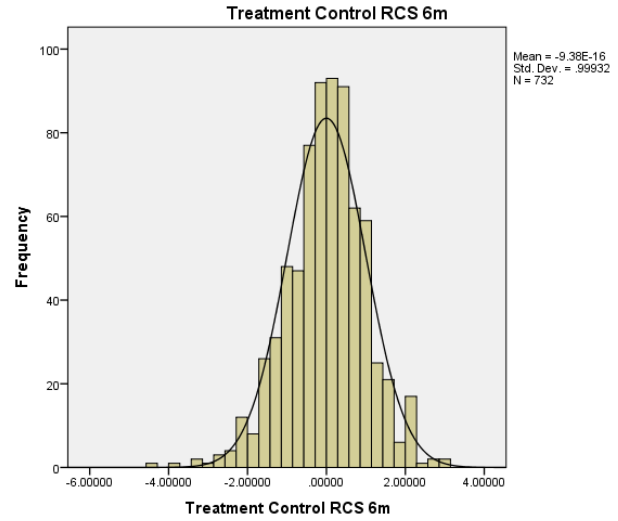

AcuteChronic RCS 6m
Mean = -1.25E-16
Std. Dev. = .99932
N = 736
Frequency
AcuteChronic RCS 6m

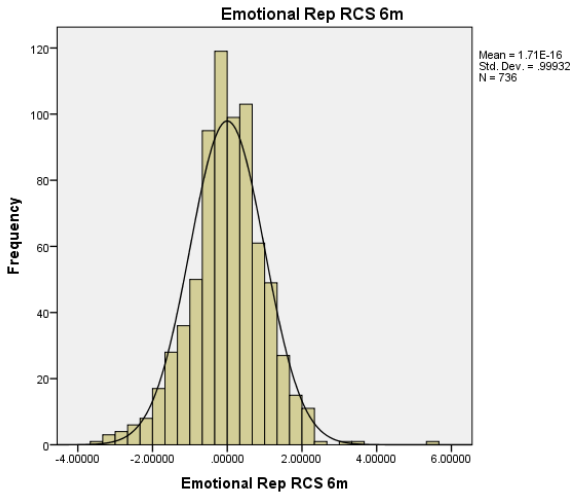

Identity RCS 6m
Mean = 7.03E-17
Std. Dev. = .99933
N = 746
Frequency
Identity RCS 6m

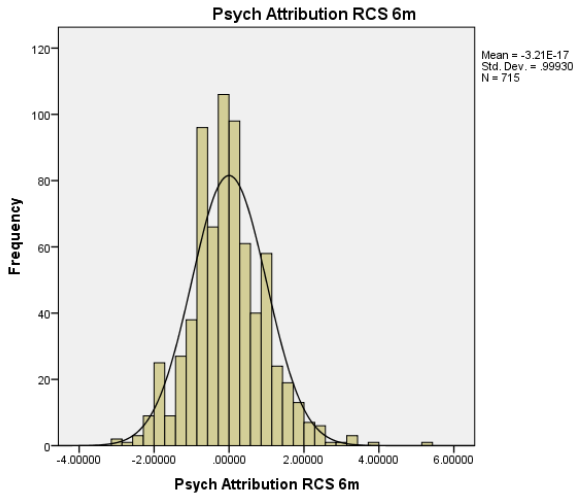

Consequences RCS 6m
Mean = 9.85E-16
Std. Dev. = .99932
N = 737
Frequency
Consequences RCS 6m

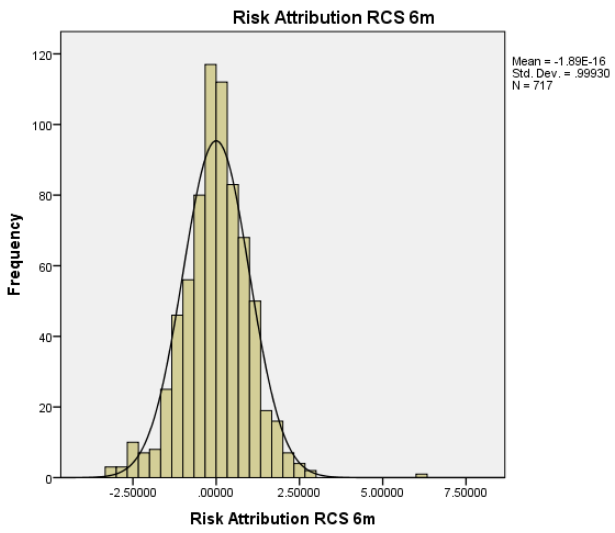

Cyclical RCS 6m
Mean = -5.53E-16
Std. Dev. = .99932
N = 733
Frequency
Cyclical RCS 6m

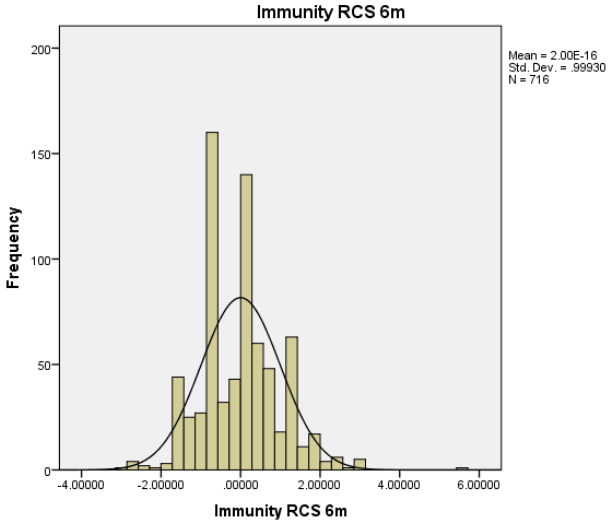

Illness Coherence RCS 6m
Mean = -1.57E-16
Std. Dev. = .99932
N = 734
Frequency
Illness Coherence RCS 6m

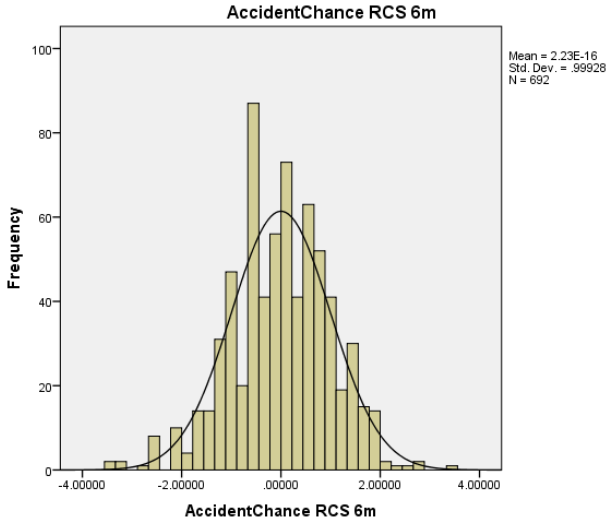

Treatment Control RCS 6m
Mean = -9.38E-16
Std. Dev. = .99932
N = 732
Frequency
Treatment Control RCS 6m

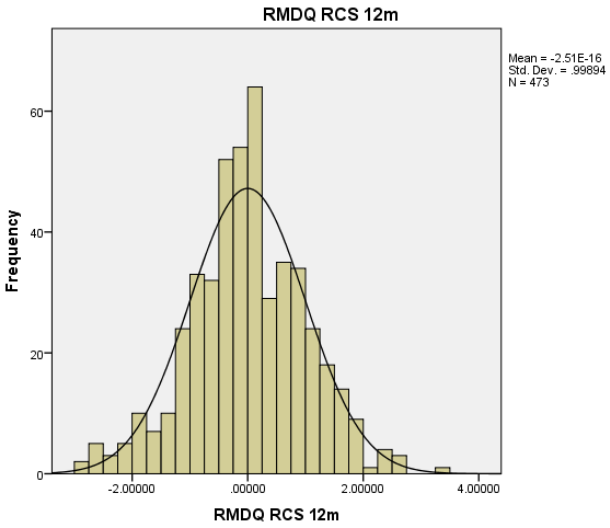

Emotional Rep RCS 6m
Mean = 1.71E-16
Std. Dev. = .99932
N = 736
Frequency
Emotional Rep RCS 6m

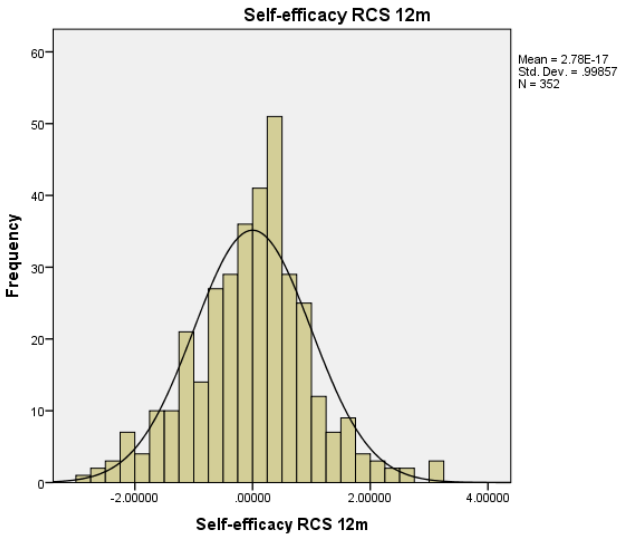

Psych Attribution RCS 6m
Mean = -3.21E-17
Std. Dev. = .99930
N = 715
Frequency
Psych Attribution RCS 6m

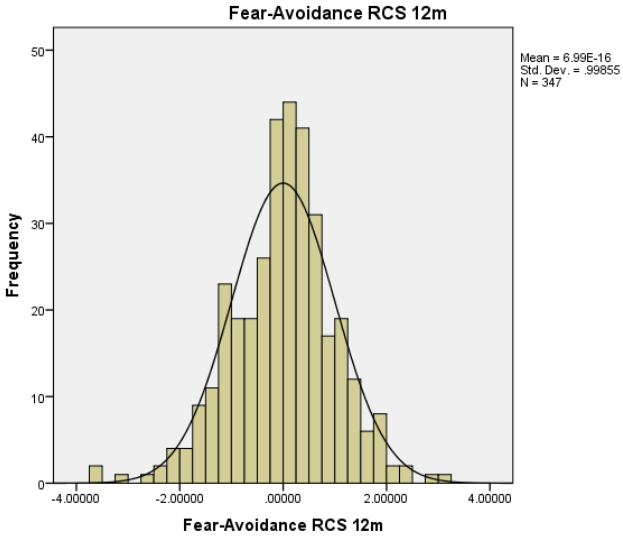

Risk Attribution RCS 6m
Mean = -1.89E-16
Std. Dev. = .99930
N = 717
Frequency
Risk Attribution RCS 6m

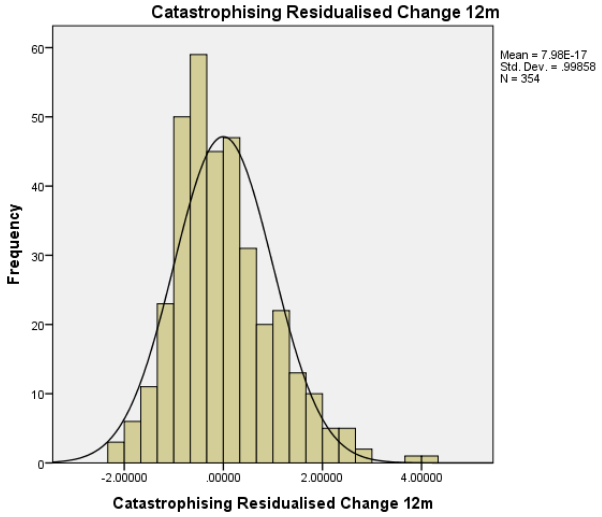

Immunity RCS 6m
Mean = 2.00E-16
Std. Dev. = .99930
N = 716
Frequency
Immunity RCS 6m

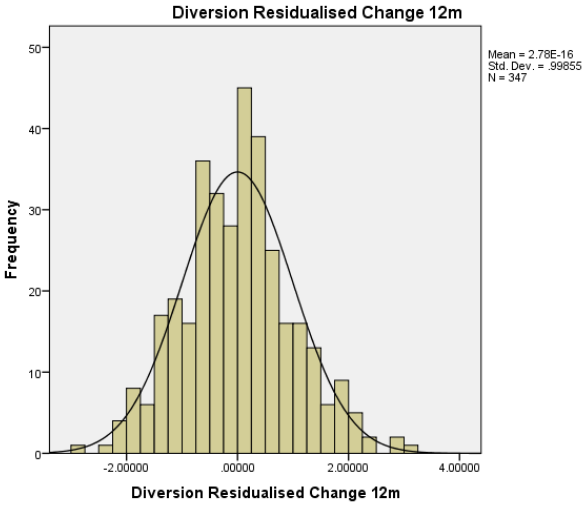

AccidentChance RCS 6m
Mean = 2.23E-16
Std. Dev. = .99928
N = 692
Frequency
AccidentChance RCS 6m

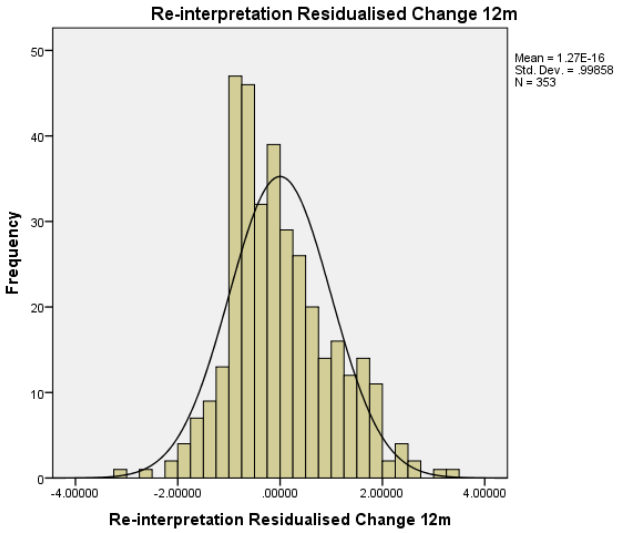

RMDQ RCS 12m
Mean = -2.51E-16
Std. Dev. = .99894
N = 473
Frequency
RMDQ RCS 12m

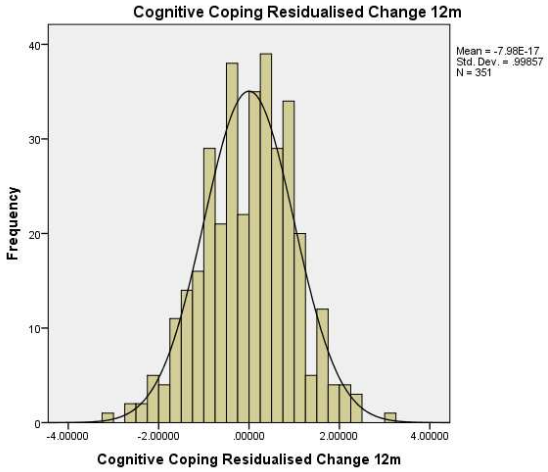

Self-efficacy RCS 12m
Mean = 2.78E-17
Std. Dev. = .99857
N = 352
Frequency
Self-efficacy RCS 12m

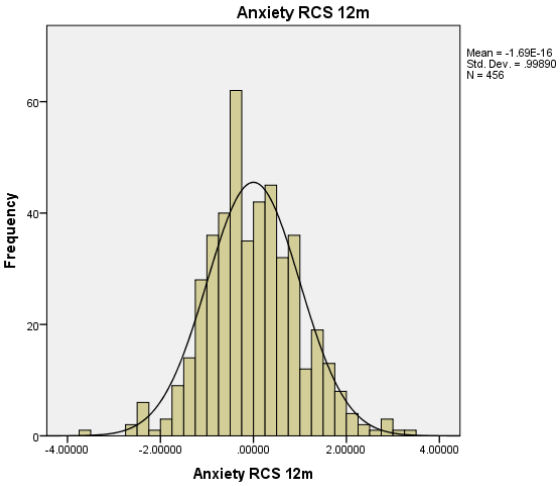

Fear-Avoidance RCS 12m
Mean = 6.99E-16
Std. Dev. = .99855
N = 347
Frequency
Fear-Avoidance RCS 12m

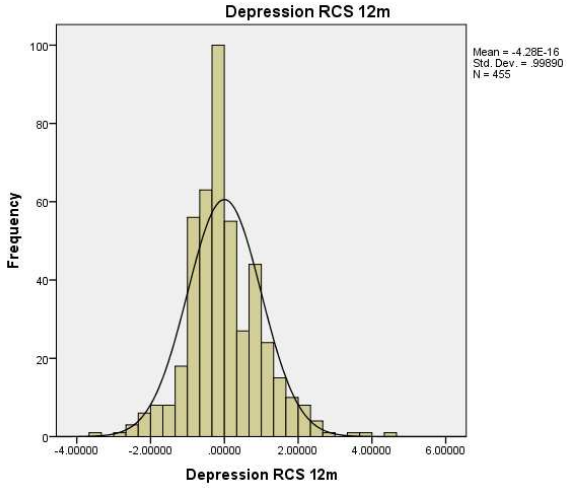

Catastrophising Residualised Change 12m
Mean = 7.98E-17
Std. Dev. = .99858
N = 354
Frequency
Catastrophising Residualised Change 12m

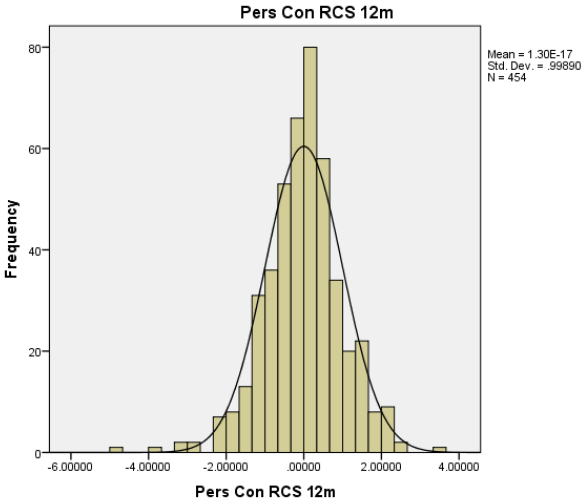

Diversion Residualised Change 12m
Mean = 2.78E-16
Std. Dev. = .99855
N = 347
Frequency
Diversion Residualised Change 12m

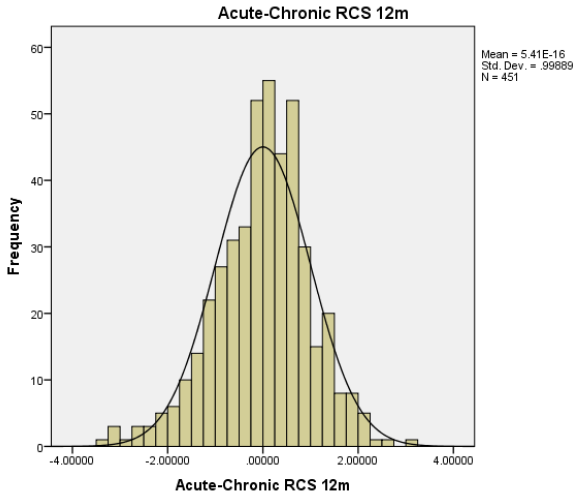

Re-interpretation Residualised Change 12m
Mean = 1.27E-16
Std. Dev. = .99858
N = 353
Frequency
Re-interpretation Residualised Change 12m

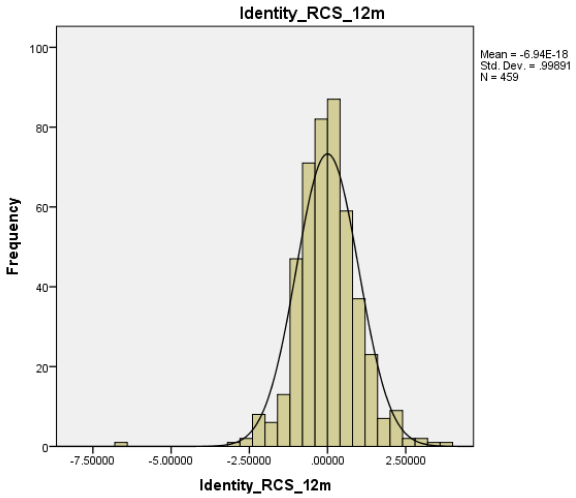

Cognitive Coping Residualised Change 12m
Mean = -7.98E-17
Std. Dev. = .99857
N = 351
Frequency
Cognitive Coping Residualised Change 12m

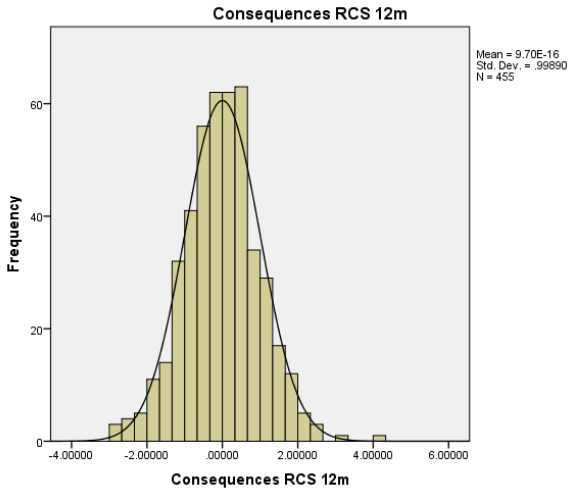

Anxiety RCS 12m
Mean = -1.69E-16
Std. Dev. = .99890
N = 456
Frequency
Anxiety RCS 12m

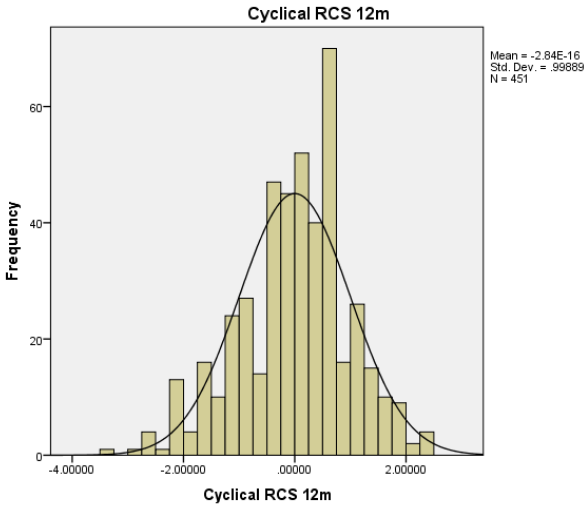

Depression RCS 12m
Mean = -4.28E-16
Std. Dev. = .99890
N = 455
Frequency
Depression RCS 12m

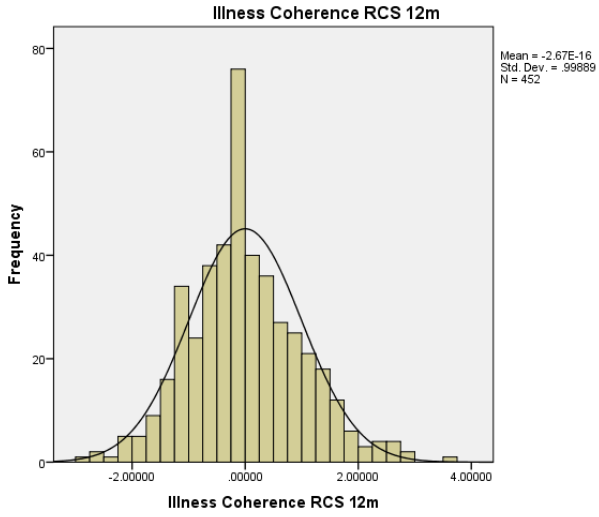

Pers Con RCS 12m
Mean = 1.30E-17
Std. Dev. = .99890
N = 454
Frequency
Pers Con RCS 12m

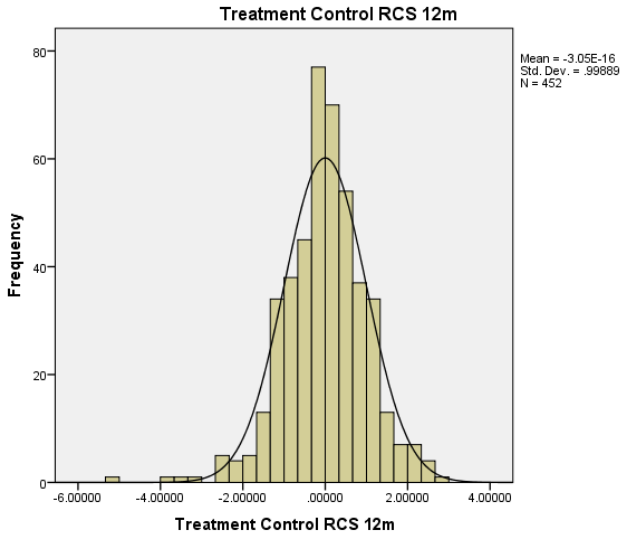

Acute-Chronic RCS 12m
Mean = 5.41E-16
Std. Dev. = .99889
N = 451
Frequency
Acute-Chronic RCS 12m

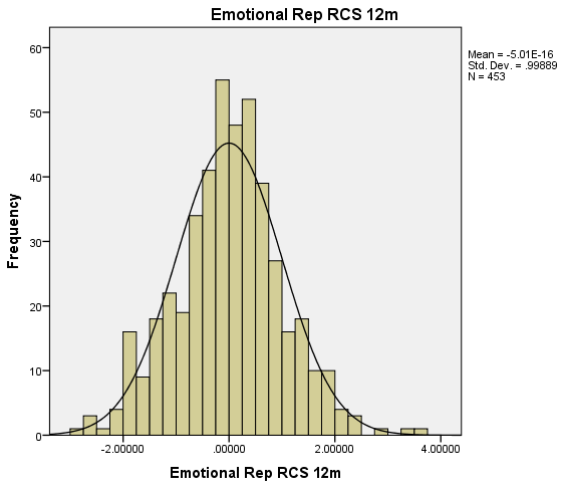

Identity_RCS_12m
Mean = -6.94E-18
Std. Dev. = .99891
N = 459
Frequency
Identity_RCS_12m

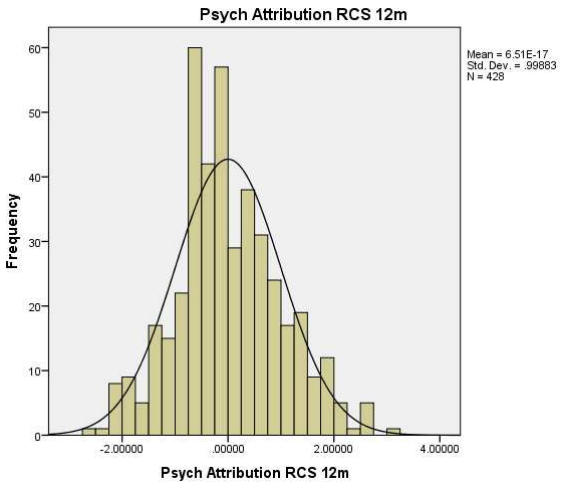

Consequences RCS 12m
Mean = 9.70E-16
Std. Dev. = .99890
N = 455
Frequency
Consequences RCS 12m

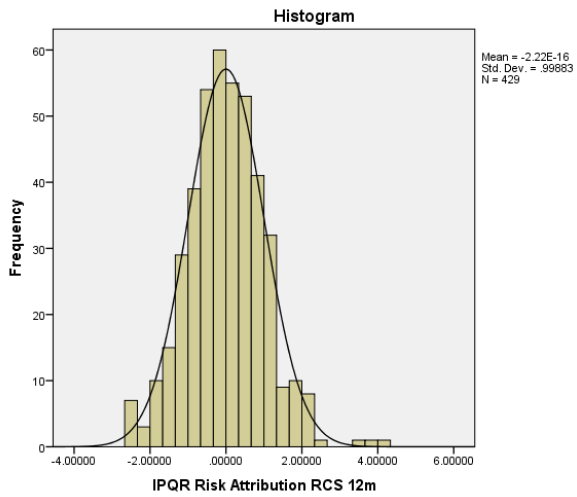

Cyclical RCS 12m
Mean = -2.84E-16
Std. Dev. = .99889
N = 451
Frequency
Cyclical RCS 12m

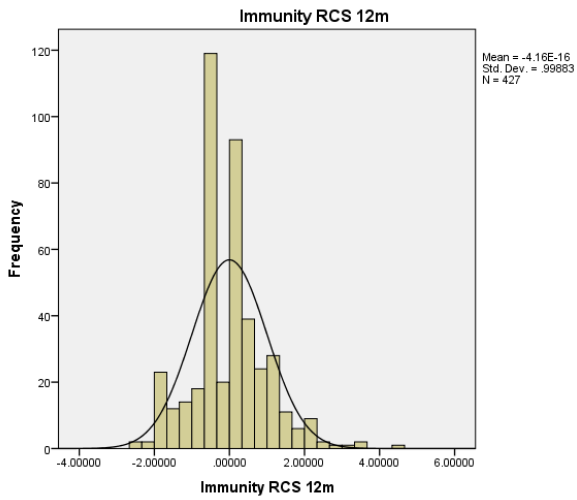

Illness Coherence RCS 12m
Mean = -2.67E-16
Std. Dev. = .99889
N = 452
Frequency
Illness Coherence RCS 12m

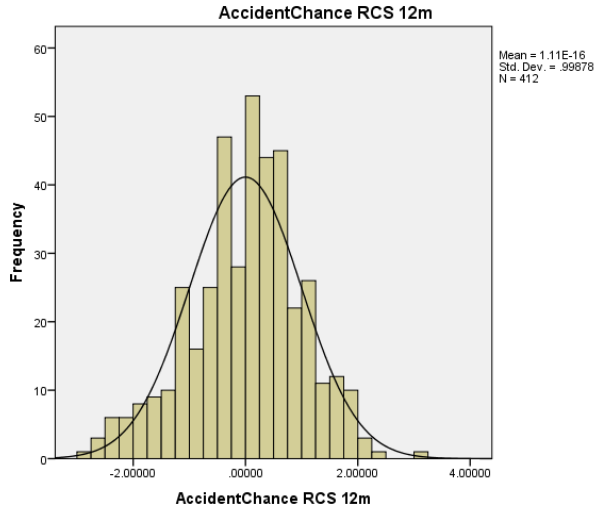

Treatment Control RCS 12m
Mean = -3.05E-16
Std. Dev. = .99889
N = 452
Frequency
Treatment Control RCS 12m


Emotional Rep RCS 12m
Mean = -5.01E-16
Std. Dev. = .99889
N = 453
Frequency
Emotional Rep RCS 12m


Psych Attribution RCS 12m
Mean = 6.51E-17
Std. Dev. = .99883
N = 428
Frequency
Psych Attribution RCS 12m


Histogram
Mean = -2.22E-16
Std. Dev. = .99883
N = 429
Frequency
IPQR Risk Attribution RCS 12m


Immunity RCS 12m
Mean = -4.16E-16
Std. Dev. = .99883
N = 427
Frequency
Immunity RCS 12m


AccidentChance RCS 12m
Mean = 1.11E-16
Std. Dev. = .99878
N = 412
Frequency
AccidentChance RCS 12m

Appendix 6.1 Results for SF-12 PCS Outcome (Complete case analysis)

Summary Details of Outcome and Predictor Measures

		Reference	**How Scored**
Outcome measure	General physical function (SF-12 PCS)	Ware et al (1996)	Higher score indicates higher level of interference

Unadjusted function outcomes at baseline and change by four-month follow-up (Means and SDs) in the STarT Back study

Variable (scale range)	Baseline means (SDs)						Mean (SD) change at four-month follow-up					
	Treatment (*n*=370)			Control (*n*=179)			Treatment (*n*=370)			Control (*n*=179)		
	L (*n*=100)	M (*n*=177)	H (*n*=93)	L (*n*=50)	M (*n*=84)	H (*n*=45)	L (*n*=100)	M (*n*=177)	H (*n*=93)	L (*n*=50)	M (*n*=84)	H (*n*=45)
SF-12 PCS (0-12)	46.86 (8.01)	34.91 (9.92)	31.22 (7.50)	45.96 (9.11)	35.94 (8.43)	30.28 (8.35)	-2.35 (8.48)	-8.78 (10.22)	-10.05 (10.19)	-1.97 (8.60)	-6.10 (9.16)	-5.16 (9.95)

Correlation coefficients with potential confounding variables

Outcome	Variable	Treatment Allocation	4m		
			Low	**Medium**	**High**
SF-12 PCS$^{\Delta}$	Age	Treatment	-0.14	-0.18	-0.18
	Sex		0.00	0.05	0.00
	Pain duration		-0.10	-0.14	-0.24*
	Age	Control	-0.01	-0.07	-0.27
	Sex		-0.08	0.19	0.01
	Pain duration		-0.12	-0.12	-0.09

*p<0.05; **p<0.01*

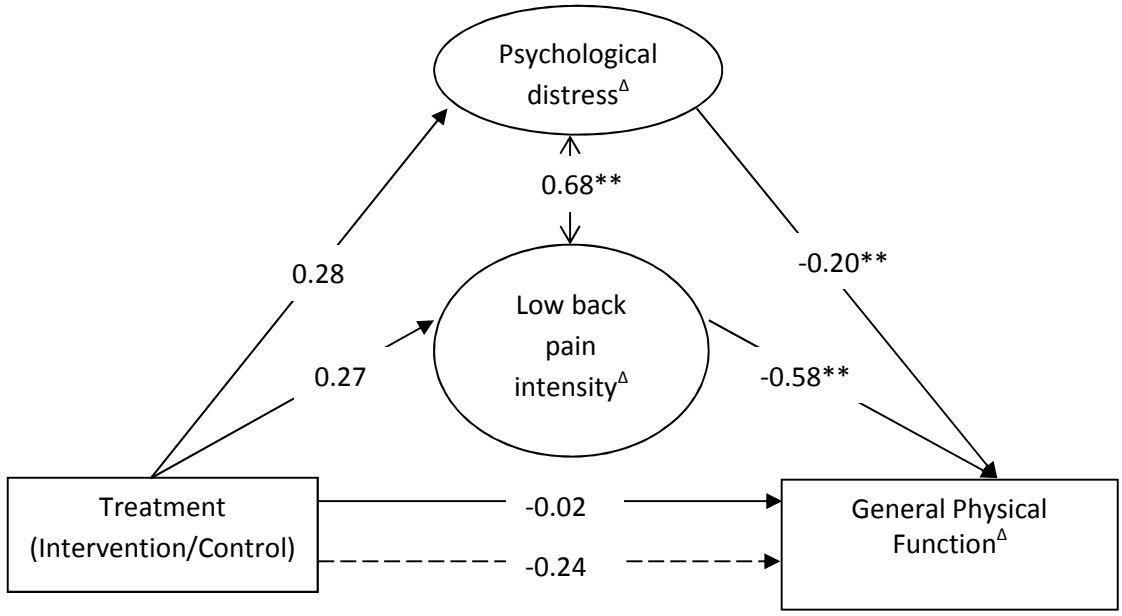
For residualised change in SF-12 PCS the same pattern emerged, with the amount of variance explained by each variable almost identical in the treatment group but less in the control group (Table 9).

Contributions of changes in each potential mediator to change in functional outcomes in STarT Back participants: linear regression analyses – High risk group

Outcome	Predictor	Treatment Allocation	Change at 4m			
			Unstandardised B (SE)	**95% CI**	**Standardised β**	**R-square change**
SF-12 PCS$^{\Delta}$	Catastrophising thoughts$^{\Delta}$	Treatment	-0.50 (0.08)	-0.67 to -0.34	-0.54	0.29
	Fear-avoidance beliefs$^{\Delta}$		-0.55 (0.80)	-0.71 to -0.39	-0.58	0.34
	Anxiety$^{\Delta}$		-0.50 (0.90)	-0.68 to -0.32	-0.50	0.25
	Depression$^{\Delta}$		-0.56 (0.09)	-0.73 to -0.39	-0.56	0.32
	Least pain$^{\Delta}$		-0.63 (0.08)	-0.79 to -0.46	-.61	0.38
	Average pain$^{\Delta}$		-0.62 (0.08)	-0.78 to -0.46	-0.63	0.40
	Current pain$^{\Delta}$		-0.67 (0.08)	-0.83 to -0.52	-0.67	0.45
	Catastrophising thoughts$^{\Delta}$	Control	-0.24 (0.17)	-0.57 to -0.10	-0.21	0.05
	Fear-avoidance beliefs$^{\Delta}$		-0.57 (0.18)	-0.94 to -0.20	-0.43	0.19
	Anxiety$^{\Delta}$		-0.21 (0.14)	-0.50 to -0.07	-0.22	0.05
	Depression$^{\Delta}$		-0.37 (0.14)	-0.65 to -0.08	-0.37	0.14
	Least pain$^{\Delta}$		-0.55 (0.12)	-0.79 to -0.32	-0.59	0.34
	Average pain$^{\Delta}$		0.80 (0.11)	-1.02 to -0.57	-0.74	0.55
	Current pain$^{\Delta}$		-0.57 (0.13)	-0.82 to -0.31	-0.56	0.32

$^{\Delta}$*Residualised change scores*

Full SEM model for mediating effect of change in psychological distress on change in general physical function: High risk group



$^{\Delta}$*residualised change*
**p<0.05*
***p<0.01*

All values are standardised

Total, Direct and Indirect Effects of each Potential Mediator on Change in General Physical Function for High Risk Patients

	Effect	**Model**	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
SF-12 PCS$^{\Delta}$	Total	-0.24 (-0.39 to -0.06)	-0.51 (-0.85 to -0.13)
	Direct	-0.03 (-0.16 to 0.09)	-0.05 (-0.34 to 0.20)
	Indirect	-0.21 (-0.34 to -0.08)	-0.45 (-0.73 to -0.16)

$^{\Delta}$*residualised change*

Model Fit Statistics for Mediation Model of Change in General Physical Function for High Risk Patients

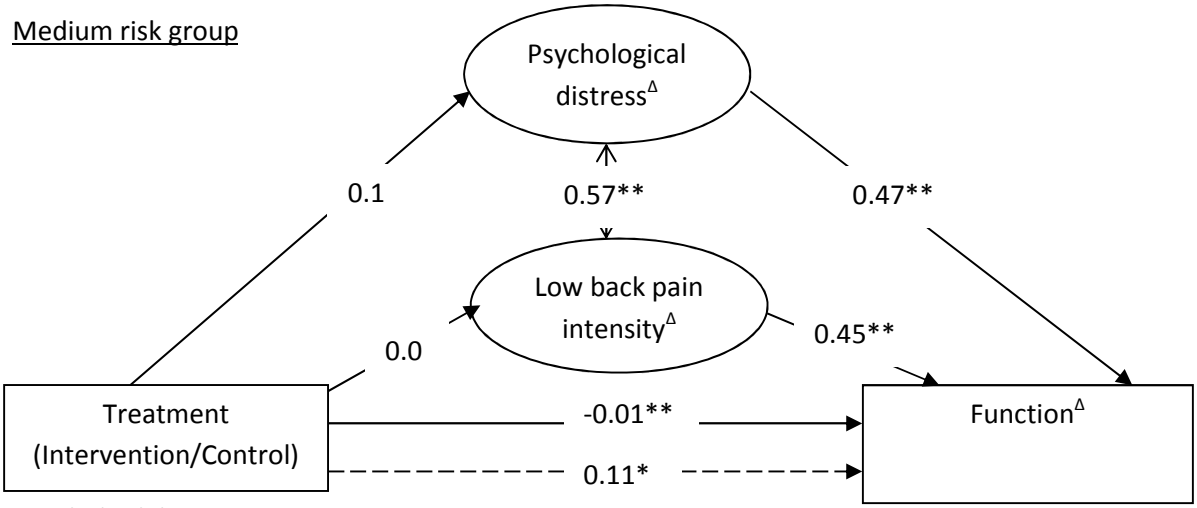
Model Index	Model	Good Model Fit
CMIN*	55.57	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.42	Between 2-5
CFI	0.96	>0.95
RMSEA	0.10 (0.07 to 0.14), PCLOSE 0.01	<0.08
SRMR	0.06	<0.08

Appendix 6.2 Results for Medium and Low Risk Groups – Complete case analysis

The model was then repeated using data for participants in the medium risk group. In this group, the percentage of variance explained by the mediation model was still high, explaining 70% of change in function overall. The pattern of correlations for change in disability in the high risk group was replicated with low, non-significant (p<0.05) beta values between treatment allocation and change in psychological distress and low back pain intensity. The correlations between change in psychological distress and change in disability score were slightly stronger in this group compared with the high-risk group, but weaker between change in low back pain intensity and change in disability. There is a weaker, but still statistically significant mediating effect of change in either latent variable on change in disability score than that seen in the high-risk models (standardised indirect effect 0.13 (95% CI 0.01 to 0.23)), and the model fit statistics also indicate a good fit of the model to the data. The weaker effect seen here was expected, as it was hypothesised that the stronger mediating effects would be in the high-risk group models where participants in the treatment arm received treatment to specifically target complexity.

Full SEM model for mediating effect of change in Psychological Distress on change in disability:

Medium risk group



Δresidualised change
**p<0.05*
***p<0.01*
All values are standardised

Total, Direct and Indirect Effects of each Potential Mediator on Disability Change for Medium Risk Patients

	Effect	**Model**	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ$^{\Delta}$	Total	0.11 (-0.01 to0 .22)	0.24 (-0.02 to 0.47)
	Direct	-0.01 (-0.09 to 0.07)	-0.03 (-0.19 to 0.14)
	Indirect	0.13 (0.01 to 0.23)	0.27 (0.03 to 0.49)

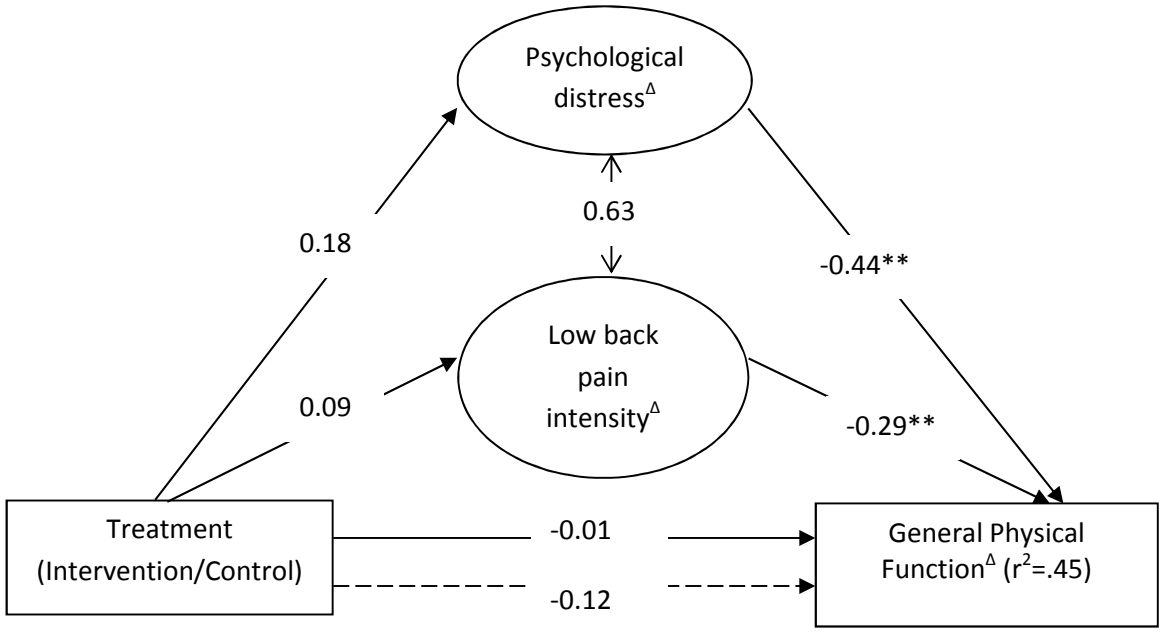
$^{\Delta}$*residualised change*

Model Fit Statistics for Mediation Model of Change in Disability Change for Medium Risk Patients

Model Index	**Model**	**Good Model Fit**
CMIN*	58.67	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.55	Between 2-5
CFI	0.97	>0.95
RMSEA	0.08 (0.05 to 0.10), PCLOSE 0.03	<0.08
SRMR	0.04	<0.08

The model for change in general physical function for the medium risk patients also presented a similar picture to the model for high risk patients. A smaller amount of variance was explained by the model overall (r^2=45%) and a weaker but statistically significant mediating effect of change in the latent variables was found (standardised indirect effect -0.11, 95% CI -0.19 to -0.01). Like the change in disability model for the medium risk group, the correlations between change in the latent variables and change in outcome were higher than the correlations between treatment allocation and the latent variables. The model fit statistics indicated an adequate fit to the data.

Figure 12 Full SEM model for mediating effect of change in psychological distress on change in general physical function: Medium risk group



Δresidualised change
**p<0.05*
***p<0.01*

All values are standardised

Total, Direct and Indirect Effects of each Potential Mediator on General Physical Function Change for Medium Risk Patients

	Effect	Model	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ$^{\Delta}$	Total	-0.12 (-0.23 to 0.02)	-0.25 (-0.49 to 0.04)
	Direct	-0.01 (-0.11 to 0.09)	-0.02 (-0.23 to 0.18)
	Indirect	-0.11 (-0.19 to -0.01)	-0.22 (-0.42 to -0.03)

$^{\Delta}$*residualised change*

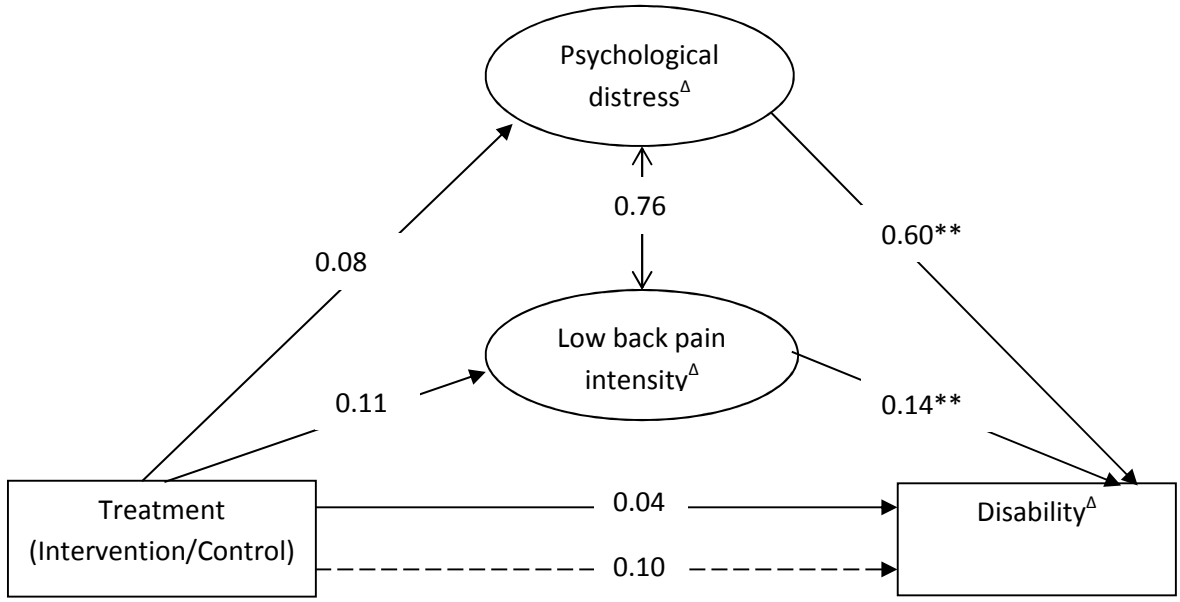
Model Fit Statistics for Mediation Model of Change in General Physical Function for Medium Risk Patients

Model Index	**Model**	**Good Model Fit**
CMIN*	49.66	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.16	Between 2-5
CFI	0.97	>0.95
RMSEA	0.07 (0.04 to 0.09), PCLOSE 0.13	<0.08
SRMR	0.04	<0.08

The model for the low risk group with change in disability as an outcome is shown below. This model explained the lowest amount of variance of any of the disability change models (51%). The indirect effect was also weakest out of all the disability models and not statistically significant (standardised indirect effect 0.06, 95% CI -0.08 to 0.21). The statistics for model fit suggested adequate fit of the model to the data.

Full SEM model for mediating effect of change in psychological distress on change in disability:

Low risk group



Δresidualised change
**p<0.05*
***p<0.01*

All values are standardised

Total, Direct and Indirect Effects of each Potential Mediator on Change in Psychological Distress for Low Risk Patients

	Effect	**Model**	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ$^{\Delta}$	Total	0.10 (-0.05 to 0.27)	0.21 (0.10 to 0.60)
	Direct	0.04 (-0.09 to 0.17)	0.08 (-0.19 to 0.36)
	Indirect	0.06 (-0.08 to 0.21)	0.13 (-0.14 to 0.47)

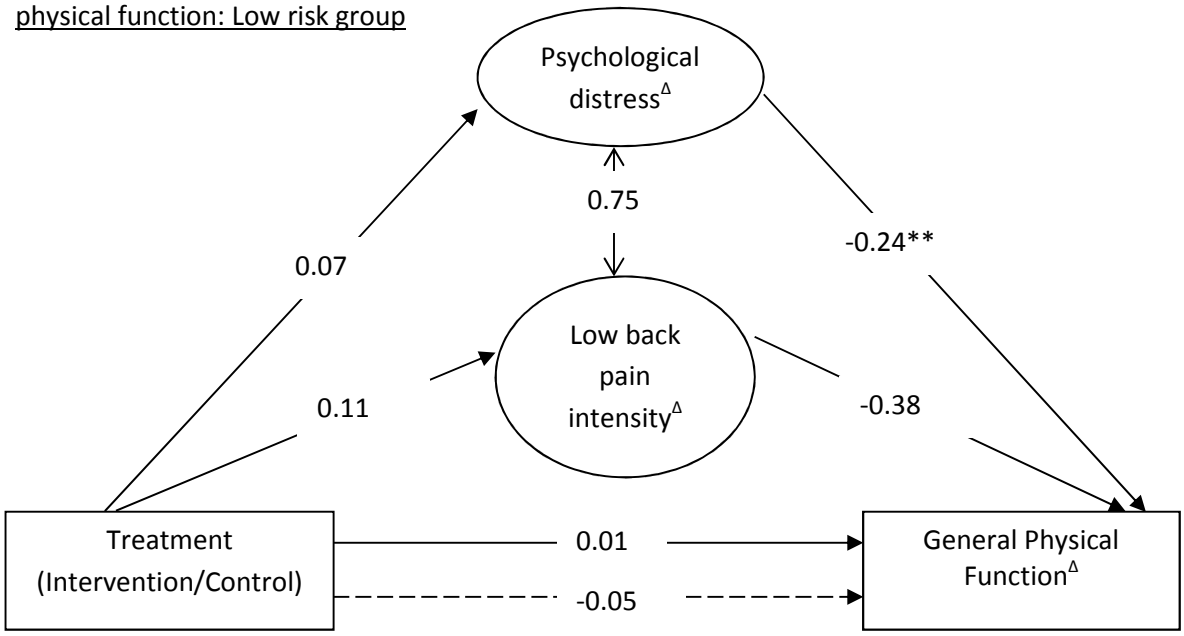
$^{\Delta}$*residualised change*

Model Fit Statistics for Mediation Model of Change Psychological Distress for Low Risk Patients

Model Index	**Model**	**Good Model Fit**
CMIN*	54.95	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.39	Between 2-5
CFI	0.94	>0.95
RMSEA	0.10 (0.06 to 0.13), PCLOSE 0.01	<0.08
SRMR	0.06	<0.08

The model for change in general physical function again showed no statistically significant mediating effect of change in the latent variables on change in SF-12 PCS score (standardised indirect effect -0.06, 95% CI -0.18 to 0.05). Like for the other models, this model provided an adequate fit to the data. The model as a whole also only explained 34% of the variance of change in general physical health, the lowest out of all the models with this outcome.

Full SEM model for mediating effect of change in psychological distress on change in general physical function: Low risk group



Δresidualised change
**p<0.05*
***p<0.01*

All values are standardised

Table 21a Total, Direct and Indirect Effects of each Potential Mediator on General Physical Function Change for Low Risk Patients

	Effect	**Model**	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ$^{\Delta}$	Total	-0.05 (-0.22 to 0.10)	-0.11 (-0.49 to 0.20)
	Direct	0.01 (-0.13 to 0.14)	0.02 (-0.27 to 0.30)
	Indirect	-0.06 (-0.18 to 0.05)	-0.12 (-0.42 to 0.09)

$^{\Delta}$*residualised change*

Table 21b Model Fit Statistics for Mediation Model of Change in General Physical Function for Low Risk Patients

Model Index	Model	Good Model Fit
CMIN*	47.93	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.08	Between 2-5
CFI	0.95	>0.95
RMSEA	0.09 (0.05 to 0.12), PCLOSE 0.05	<0.08
SRMR	0.06	<0.08

Appendix 6.3 Full analysis of High-risk Group using all available data (FIML)

Preliminary analysis

Outcome measure	Baseline score (Mean and SD)		Four-month Follow-up (Mean change and SD)	
	Treatment (*n*=157)	Control (*n*=79)	Treatment (*n*=131)	Control (*n*=56)
Disability	14.01 (4.64)	14.72 (4.40)	6.77 (6.81)	4.21 (5.03)
General physical function	30.83 (7.60)	29.57 (8.25)	-8.98 (10.36)	-5.37 (9.93)

Potential mediator	Baseline score (Mean and SD)		Four-month Follow-up (Mean change and SD)	
	Treatment (*n*=157)	Control (*n*=79)	Treatment (*n*=131)	Control (*n*=56)
Catastrophising thoughts	26.41 (10.63)	26.89 (10.27)	10.33 (12.36)	6.82 (10.41)
Fear-avoidance beliefs	45.81 (5.04)	45.96 (5.65)	8.55 (7.44)	3.67 (4.65)
Anxiety	10.11 (4.15)	10.09 (3.77)	3.00 (3.95)	2.46 (3.90)
Depression	8.90 (4.32)	8.91 (3.68)	3.13 (3.96)	1.56 (3.38)
Least Pain	6.33 (2.56)	5.84 (3.04)	2.96 (2.87)	1.88 (3.29)
Average Pain	7.94 (1.93)	8.05 (1.76)	3.87 (3.20)	2.51 (2.93)
Current Pain	6.68 (2.20)	6.58 (2.40)	2.81 (2.81)	1.47 (2.89)

Correlation Analysis – Potential Confounders

Outcome	Variable	Treatment Allocation	4 months
			High-risk group
RMDQ$^{\Delta}$	Age	Treatment	0.08
	Sex		-0.02
	Pain duration		0.32**
	Age	Control	-0.00
	Sex		0.22
	Pain duration		0.35**

Testing Criteria for Mediation – Linear regressions to show associations between change in potential mediator and change in outcome (b path)

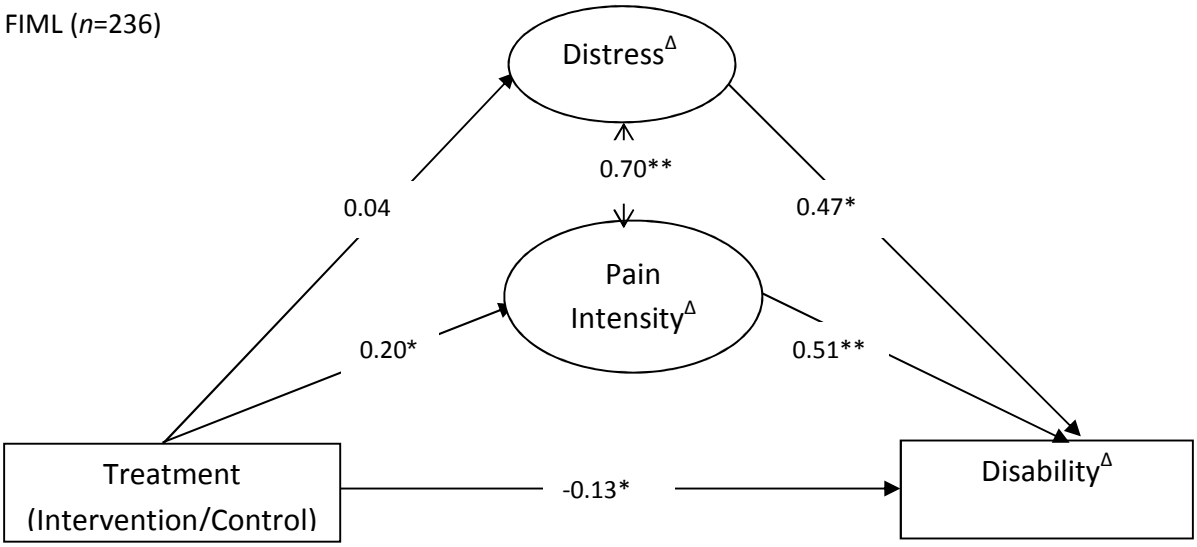
Outcome	Predictor	Treatment Allocation	Change at 4 months			
			Unstandardised B (SE)	95% CI	Standardised B	R-square change
RMDQ$^{\Delta}$	Catastrophising thoughts$^{\Delta}$	Treatment	0.54 (0.07)	0.40 to 0.67	0.59	0.34
	Fear-avoidance beliefs$^{\Delta}$		0.58 (0.08)	0.42 to 0.74	0.57	0.32
	Anxiety$^{\Delta}$		0.56 (0.08)	0.40 to 0.73	0.55	0.30
	Depression$^{\Delta}$		0.58 (0.08)	0.43 to 0.73	0.59	0.35
	Least pain$^{\Delta}$		0.86 (0.09)	0.69 to 1.03	0.69	0.47
	Average pain$^{\Delta}$		0.88 (0.08)	0.72 to 1.04	0.72	0.52
	Current pain$^{\Delta}$		1.00 (0.07)	0.85 to 1.14	0.79	0.62
	Catastrophising thoughts$^{\Delta}$	Control	0.53 (0.09)	0.35 to 0.71	0.64	0.41
	Fear-avoidance beliefs$^{\Delta}$		0.63 (0.13)	0.37 to 0.88	0.58	0.33
	Anxiety$^{\Delta}$		0.48 (0.11)	0.27 to 0.69	0.54	0.29
	Depression$^{\Delta}$		0.57 (0.11)	0.34 to 0.80	0.58	0.34
	Least pain$^{\Delta}$		0.64 (0.12)	0.40 to 0.87	0.62	0.38
	Average pain$^{\Delta}$		0.70 (0.13)	0.45 to 0.96	0.62	0.38
	Current pain$^{\Delta}$		0.57 (0.12)	0.33 to 0.82	0.55	0.30

Testing Criteria for Mediation – Linear regressions to show associations between treatment allocation and change in potential mediator (a path)

Outcome	Predictor	Unstandardised B (SE)	95% CI	Standardised B	R-square change
Catastrophising thoughts$^{\Delta}$	Treatment Allocation	0.19 (0.22)	-0.25 to 0.63	0.07	0.00
Fear-avoidance beliefs$^{\Delta}$	Treatment Allocation	0.23 (0.19)	-0.15 to 0.60	0.10	0.01
Anxiety$^{\Delta}$	Treatment Allocation	0.23 (0.18)	-0.17 to 0.62	0.09	0.01
Depression$^{\Delta}$	Treatment Allocation	0.04 (0.20)	-0.35 to 0.43	0.02	0.00
Least pain$^{\Delta}$	Treatment Allocation	0.38 (0.17)	0.05 to 0.70	0.18	0.03
Average pain$^{\Delta}$	Treatment Allocation	0.52 (0.16)	0.20 to 0.84	0.25	0.06
Current pain$^{\Delta}$	Treatment Allocation	-.40 (0.16)	0.18 to 0.81	0.23	0.05

High-risk group

FIML (*n*=236)



$^{\Delta}$*residualised change*
**p<0.05*
***p<0.01*
All values are standardised

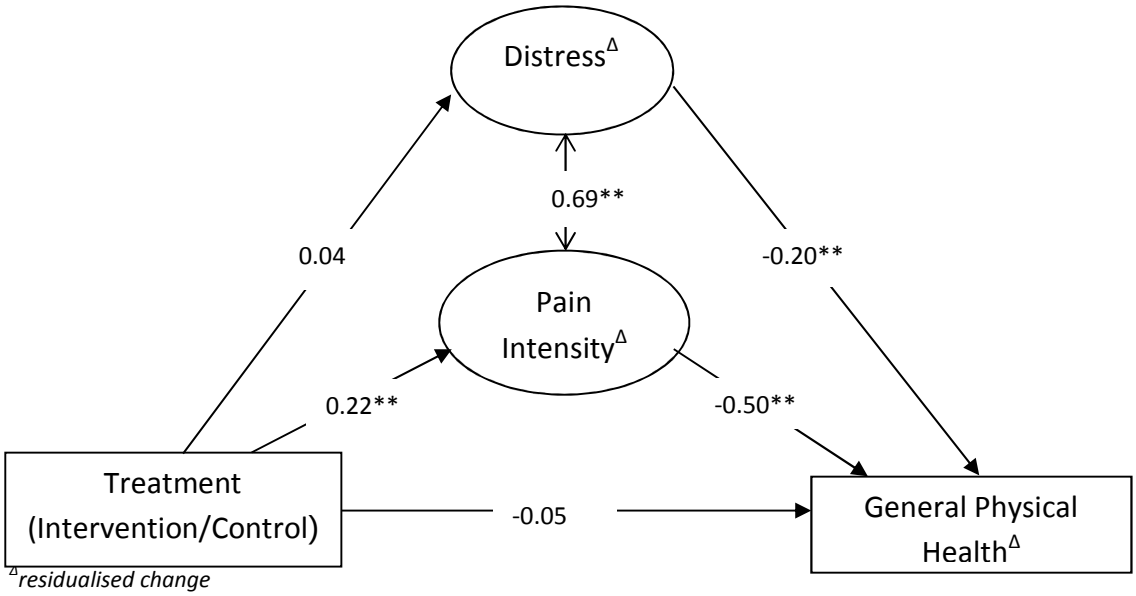
Total, Direct and Indirect Effects of each Potential Mediator on Change in Disability for High Risk Patients

	Effect	Model	
		Standardised	Unstandardised
RMDQ$^{\Delta}$	Total	-0.02	-0.04
	Direct	-0.13	-0.33
	Indirect	0.11	0.29

$^{\Delta}$*residualised change*
**p<0.05*

Model Fit Statistics for Mediation Model of Change in Disability for High Risk Patients

Model Index	Current Model	Good Model Fit
CMIN*	44.67	Non-significant result
DF	23	
P	0.00	
CMIN/DF	1.94	Between 2-5
CFI	0.97	Closer to 1
RMSEA	0.06 (0.04 to 0.09, PCLOSE 0.20)	<0.08
SRMR	-	<0.08



$^{\Delta}$*residualised change*
**p<0.05*
***p<0.01*
All values are standardised

Total, Direct and Indirect Effects of each Potential Mediator on Change in General Physical Health for High Risk Patients

	Effect	Model	
		Standardised	Unstandardised
SF-12 PCS$^{\Delta}$	Total	-0.17	-0.36
	Direct	-0.05	-0.10
	Indirect	-0.12	-0.25

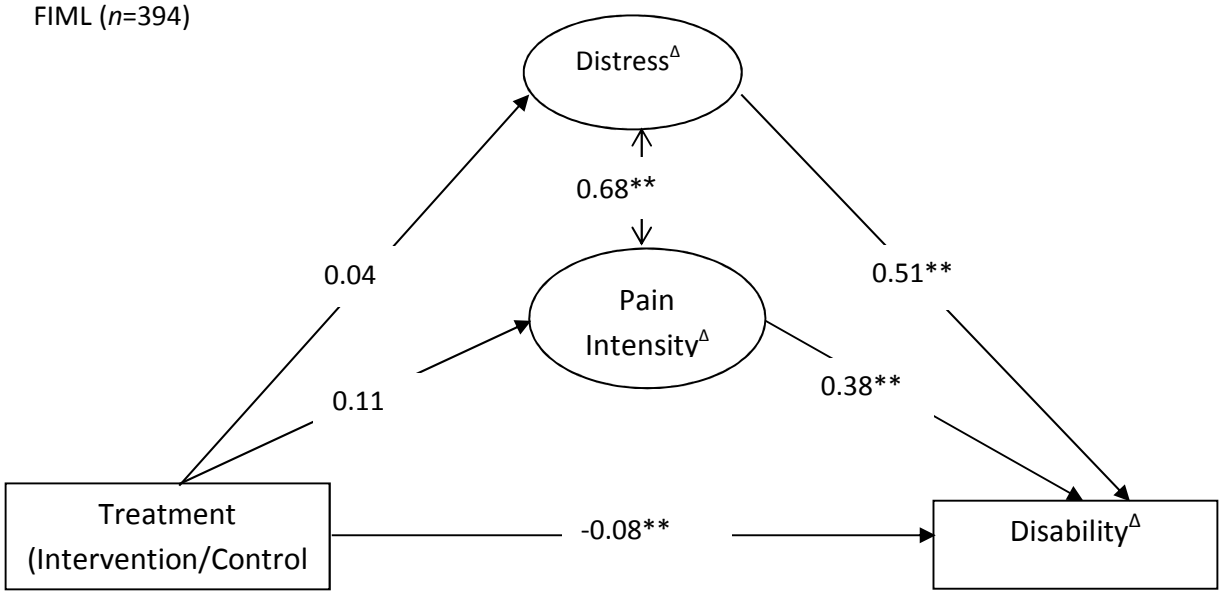
$^{\Delta}$*residualised change*
**p<0.05*

Model Fit Statistics for Mediation Model of Change in General Physical Health for High Risk Patients

Model Index	**Current Model**	**Good Model Fit**
CMIN*	47.58	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.07	Between 2-5
CFI	0.97	Closer to 1
RMSEA	0.07 (0.04 to 0.10, PCLOSE 0.17)	<0.08
SRMR	-	<0.08

Medium Risk Group

FIML (*n*=394)



Δresidualised change
**p<0.05*
***p<0.01*

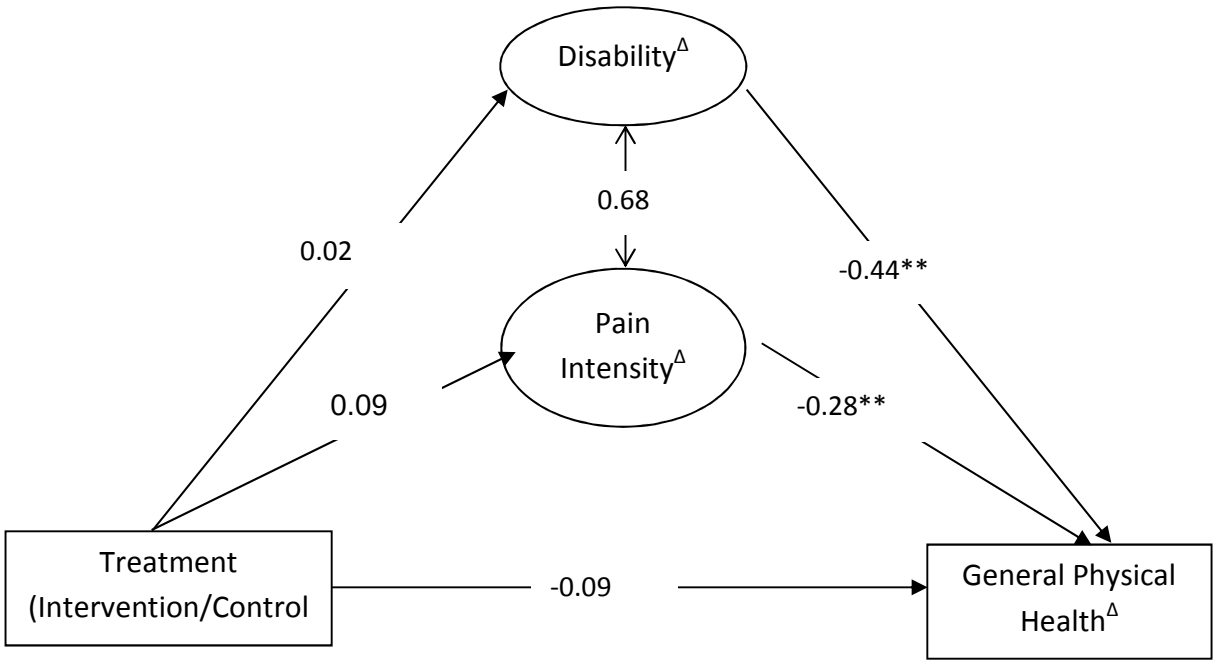
Total, Direct and Indirect Effects of each Potential Mediator on Disability Change for Medium Risk Patients

	Effect	Model	
		Standardised	Unstandardised
RMDQ$^{\Delta}$	Total	-0.02	-0.04
	Direct	-0.08	-0.17
	Indirect	0.06	0.13

$^{\Delta}$*residualised change*
**p<0.05*

Model Fit Statistics for Mediation Model of Change in Disability Change for Medium Risk Patients

Model Index	Current Model	Good Model Fit
CMIN*	78.51	Non-significant result
DF	23	
P	0.00	
CMIN/DF	3.41	Between 2-5
CFI	0.95	>0.95
RMSEA	0.08 (0.06 to 0.10, PCLOSE 0.01)	<0.08
SRMR	-	<0.08



Δresidualised change
**p<0.05*
***p<0.01*

Total, Direct and Indirect Effects of each Potential Mediator on General Physical Health Change for Medium Risk Patients

	Effect	Model	
		Standardised	Unstandardised
SF12$^{\Delta}$	Total	-0.13	-0.28
	Direct	-0.10	-0.20
	Indirect	-0.04	-0.08

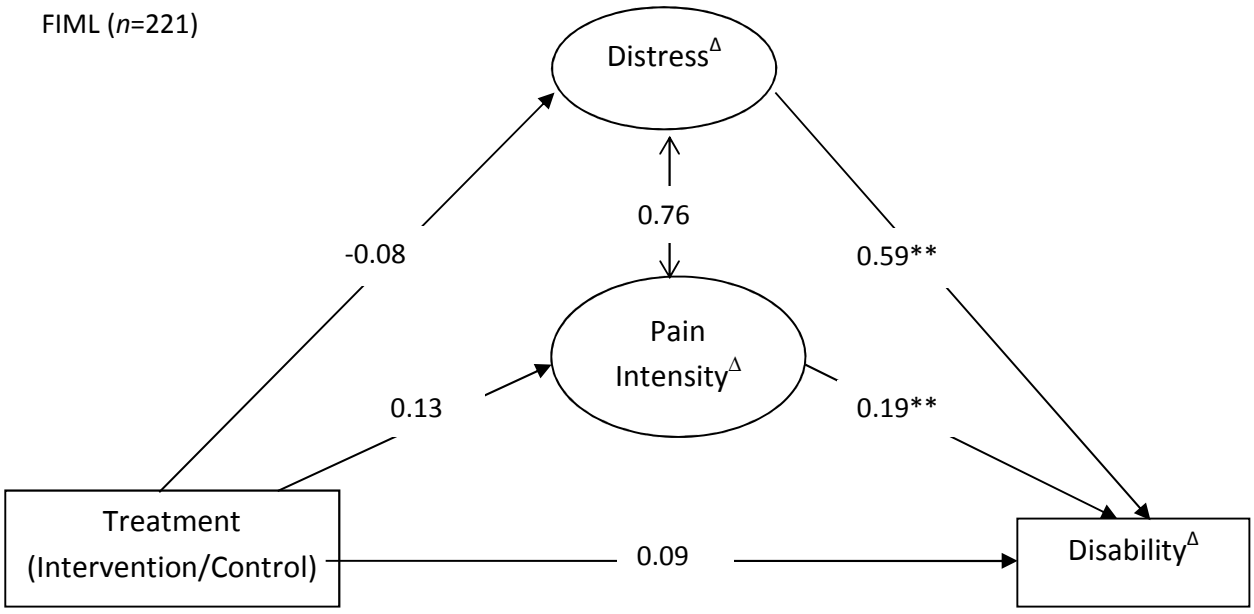
$^{\Delta}$*residualised change*
**p<0.05*

Model Fit Statistics for Mediation Model of Change in General Physical Health for Medium Risk Patients

Model Index	**Current Model**	**Good Model Fit**
CMIN*	61.14	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.66	Between 2-5
CFI	0.96	>0.95
RMSEA	0.07 (0.05 to 0.09, PCLOSE 0.10)	<0.08
SRMR	-	<0.08

Low Risk Group

FIML (*n*=221)



Δresidualised change
**p<0.05*
***p<0.01*

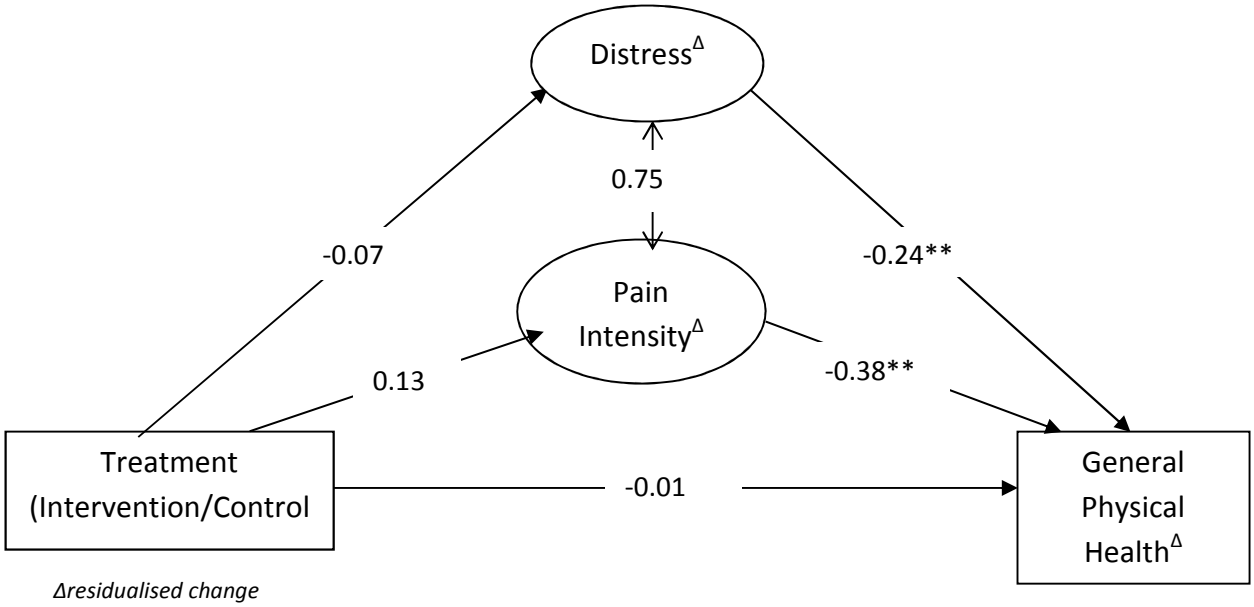
Total, Direct and Indirect Effects of each Potential Mediator on Change in Distress for Low Risk Patients

	Effect	Model	
		Standardised	Unstandardised
RMDQ$^{\Delta}$	Total	0.07	0.10
	Direct	0.09	0.13
	Indirect	-0.02	-0.03

$^{\Delta}$*residualised change*
**p<0.05*

Model Fit Statistics for Mediation Model of Change Distress for Low Risk Patients

Model Index	**Current Model**	**Good Model Fit**
CMIN*	49.86	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.17	Between 2-5
CFI	0.95	>0.95
RMSEA	0.07 (0.05 to 0.10, PCLOSE 0.08)	<0.08
SRMR	-	<0.08



Δresidualised change
**p<0.05*
***p<0.01*

Total, Direct and Indirect Effects of each Potential Mediator on General Health Change for Low Risk Patients

	Effect	**Model**	
		Standardised	Unstandardised
RMDQ$^{\Delta}$	Total	-0.05	-0.09
	Direct	-0.01	-0.03
	Indirect	-0.03	-0.07

$^{\Delta}$*residualised change*

Model Fit Statistics for Mediation Model of Change in General Health for Low Risk Patients

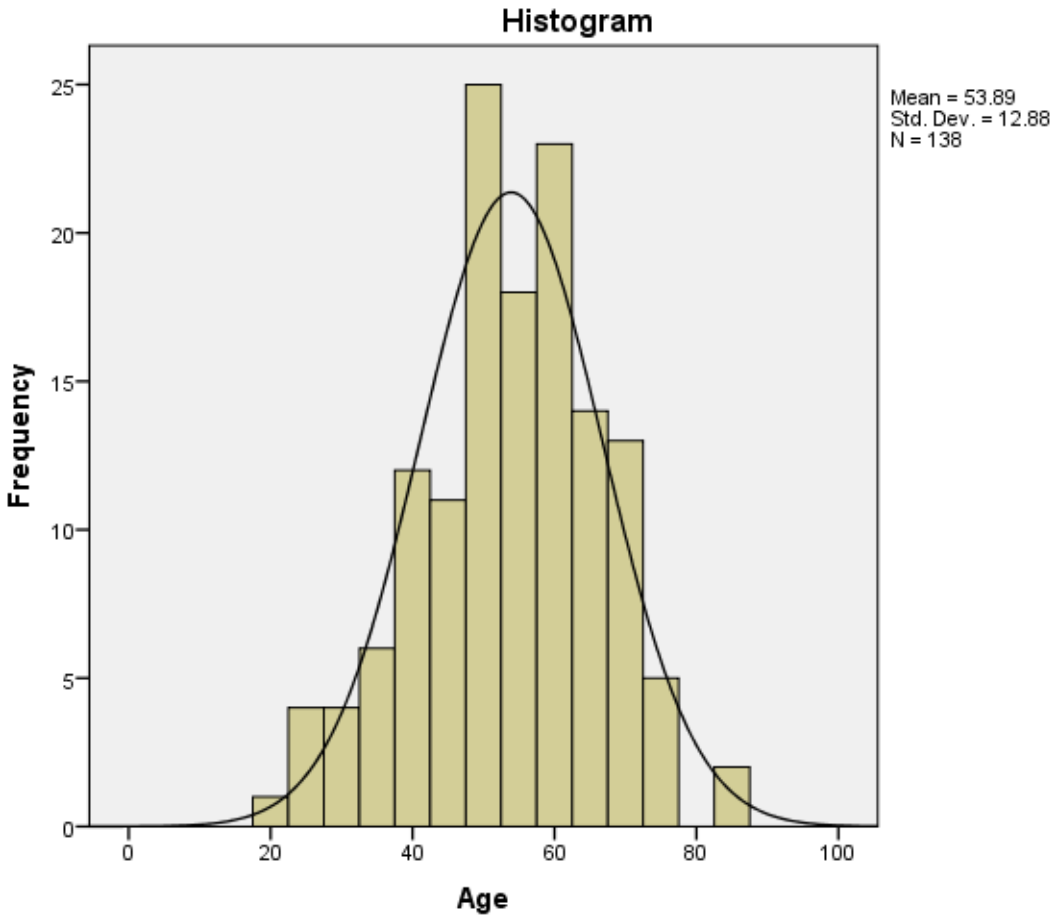
Model Index	**Current Model**	**Good Model Fit**
CMIN*	41.47	Non-significant result
DF	23	
P	0.00	
CMIN/DF	1.80	Between 2-5
CFI	0.96	>0.95
RMSEA	0.06 (0.03 to 0.09, PCLOSE 0.26)	<0.08
SRMR	-	<0.08

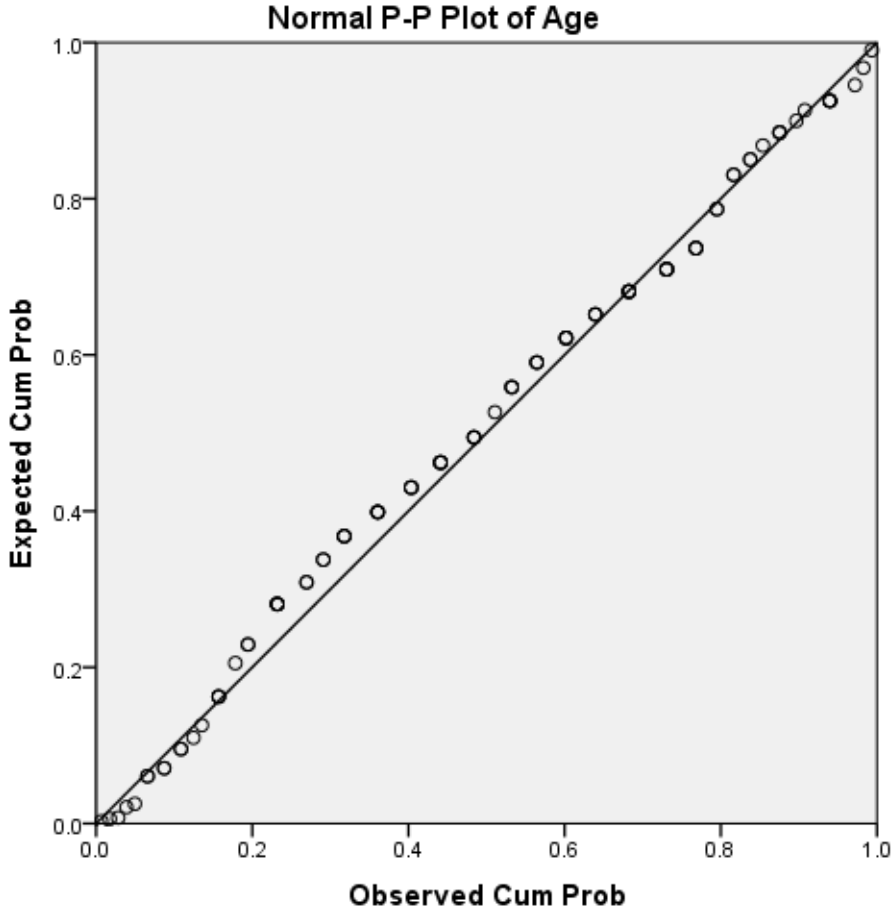
Appendix 6.4 Descriptive Statistics for Age to check for normal distribution

Statistics

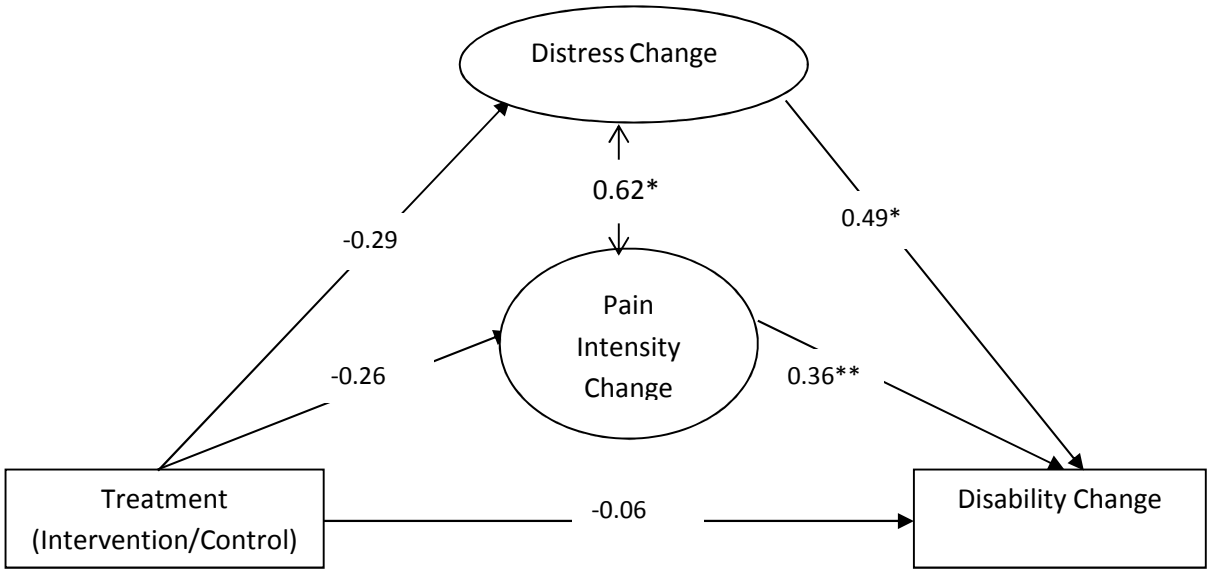
Age

N	Valid	138
	Missing	0
Mean		53.89
Std. Deviation		12.880
Skewness		-.259
Std. Error of Skewness		.206
Kurtosis		-.019
Std. Error of Kurtosis		.410




Normal P-P Plot of Age
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

Appendix 6.5 Raw Change Scores – High Risk Analysis

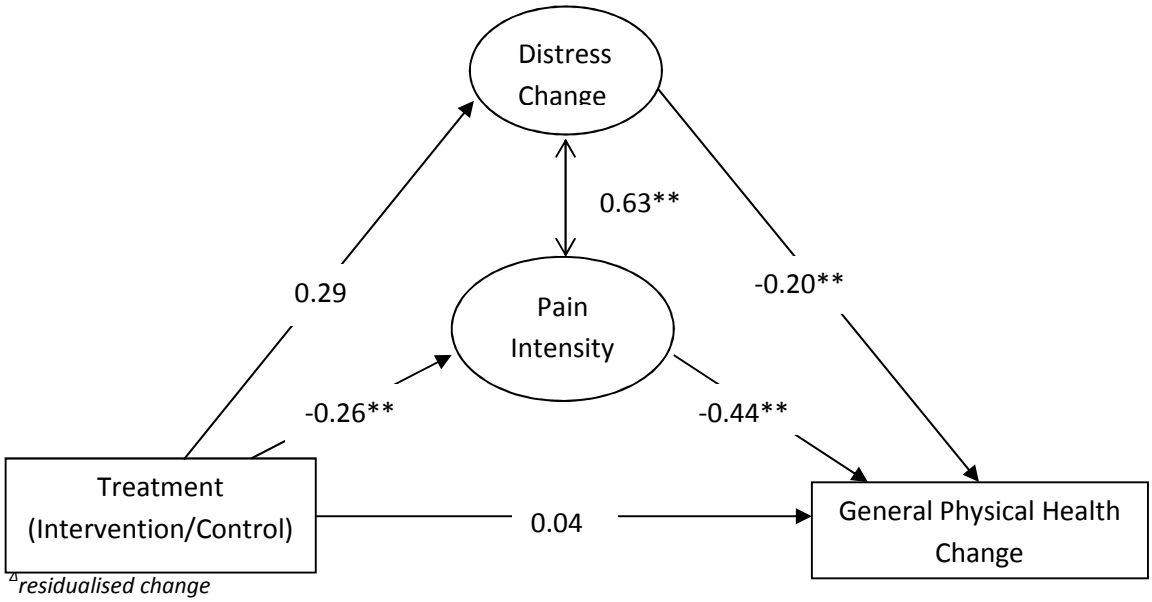


Total, Direct and Indirect Effects of each Potential Mediator on Change in Disability for High Risk Patients

	Effect	**Model**	
		Standardised (95% CI)	Unstandardised (95% CI)
RMDQ$^{\Delta}$	Total	-1.53 (-0.43 to -0.15)	-3.87 (-5.65 to -1.89)
	Direct	0.07 (-0.18 to 0.07)	0.93 (-2.30 to 0.93)
	Indirect	-0.09 (-0.39 to -0.09)	-1.09 (-5.01 to -1.09)

Model Fit Statistics for Mediation Model of Change in Disability for High Risk Patients

Model Index	**Current Model**	**Good Model Fit**
CMIN*	46.97	Non-significant result
DF	23	
P	0.00	
CMIN/DF	2.04	Between 2-5
CFI	0.95	Closer to 1
RMSEA	0.09 (0.05 to 0.12, PCLOSE 0.05)	<0.08
SRMR	0.06	<0.08



Δ residualised change
**p<0.05*

Total, Direct and Indirect Effects of each Potential Mediator on Change in Disability for High Risk Patients

	Effect	**Model**	
		Standardised (95% CI)	Unstandardised (95% CI)
SF12 PCS	Total	0.06 (0.06 to 0.37)	1.23 (1.23 to 8.28)
	Direct	0.20 (-0.11 to 0.20)	4.60 (-2.54 to 4.60)
	Indirect	0.32 (0.06 to 0.32)	7.02 (1.34 to 7.02)

Model Fit Statistics for Mediation Model of Change in Disability for High Risk Patients

Model Index	**Current Model**	**Good Model Fit**
CMIN*	51.18	Non-significant result
DF	23	
P	.00	
CMIN/DF	2.23	Between 2-5
CFI	0.94	Closer to 1
RMSEA	0.10 (0.06 to 0.13, PCLOSE 0.02)	<0.08
SRMR	0.06	<0.08

Appendix 6.6 STarT Back Analysis: Tests of normal distribution (Complete case analysis)

Baseline Scores: Treatment/Low Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		1.17 (0.24)	1.90 (0.48)
General Physical Health (SF-12 PCS)		-0.66 (0.24)	0.91 (0.48)
Catastrophising (PCS)		1.62 (0.24)	4.35 (0.48)
Fear avoidance (TSK)		0.02 (0.24)	0.73 (0.48)
Anxiety (HADS A)		0.72 (0.24)	0.50 (0.48)
Depression (HADS D)		1.12 (0.24)	0.47 (0.48)
Pain Intensity	Least	0.95 (0.24)	0.62 (0.48)
	Average	-0.06 (0.24)	-0.94 (0.48)
	Current	0.71 (0.24)	-0.06 (0.48)

Baseline Scores: Treatment/Medium Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		0.21 (0.18)	-0.63 (0.36)
General Physical Health (SF-12 PCS)		0.17 (0.18)	-0.26 (0.36)
Catastrophising (PCS)		0.67 (0.18)	0.31 (0.36)
Fear avoidance (TSK)		-0.03 (0.18)	0.78 (0.36)
Anxiety (HADS A)		0.49 (0.18)	-0.06 (0.36)
Depression (HADS D)		0.66 (0.18)	-0.34 (0.36)
Pain Intensity	Least	0.41 (0.18)	-0.51 (0.36)
	Average	-0.50 (0.18)	0.10 (0.36)
	Current	0.09 (0.18)	-0.73 (0.36)

Baseline Scores: Treatment/High Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		-0.21 (0.25)	-0.54 (0.50)
General Physical Health (SF-12 PCS)		0.11 (0.25)	-0.14 (0.50)
Catastrophising (PCS)		0.03 (0.25)	-0.34 (0.50)
Fear avoidance (TSK)		0.29 (0.25)	1.22 (0.50)
Anxiety (HADS A)		-0.02 (0.25)	-0.08 (0.50)
Depression (HADS D)		0.28 (0.25)	-0.46 (0.50)
Pain Intensity	Least	-0.02 (0.25)	-1.12 (0.50)
	Average	-0.95 (0.25)	0.78 (0.50)
	Current	-0.32 (0.25)	-0.48 (0.50)

Baseline Scores: Control/Low Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		1.20 (0.34)	1.19 (0.66)
General Physical Health (SF-12 PCS)		-0.47 (0.34)	-0.10 (0.66)
Catastrophising (PCS)		1.20 (0.34)	2.19 (0.66)
Fear avoidance (TSK)		-1.15 (0.34)	2.01 (0.66)
Anxiety (HADS A)		0.50 (0.34)	0.23 (0.66)
Depression (HADS D)		1.82 (0.34)	4.24 (0.66)
Pain Intensity	Least	0.99 (0.34)	1.18 (0.66)
	Average	0.02 (0.34)	-0.89 (0.66)
	Current	1.11 (0.34)	1.19 (0.66)

Baseline Scores: Control/Medium Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		-0.26 (0.26)	-0.84 (0.52)
General Physical Health (SF-12 PCS)		0.05 (0.26)	0.21 (0.52)
Catastrophising (PCS)		0.98 (0.26)	0.36 (0.52)
Fear avoidance (TSK)		-0.39 (0.26)	1.96 (0.52)
Anxiety (HADS A)		0.56 (0.26)	-0.42 (0.52)
Depression (HADS D)		1.44 (0.26)	2.15 (0.52)
Pain Intensity	Least	0.47 (0.26)	-0.51 (0.52)
	Average	-0.18 (0.26)	-0.76 (0.52)
	Current	0.14 (0.26)	-0.87 (0.52)

Baseline Scores: Control/High Risk

Variable		Skewness (SE)	Kurtosis (SE)
Disability (RMDQ)		-0.65 (0.35)	0.24 (0.70)
General Physical Health (SF-12 PCS)		0.67 (0.35)	-0.51 (0.70)
Catastrophising (PCS)		0.04 (0.35)	-0.51 (0.70)
Fear avoidance (TSK)		1.02 (0.35)	2.75 (0.70)
Anxiety (HADS A)		0.42 (0.35)	-0.39 (0.70)
Depression (HADS D)		0.71 (0.35)	1.84 (0.70)
Pain Intensity	Least	-0.46 (0.35)	-0.76 (0.70)
	Average	-1.18 (0.35)	0.89 (0.70)
	Current	-0.60 (0.35)	-0.45 (0.70)

Four-month follow-up scores: Treatment/Low Risk

Variable		Skewness (SE)		Kurtosis (SE)	
		Original Score	**Residualised change score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		2.22 (0.24)	1.97 (0.24)	4.84 (0.48)	4.76 (0.48)
General Physical Health (SF-12 PCS)		-0.90 (0.24)	-0.65 (0.24)	0.75 (0.48)	0.10 (0.48)
Catastrophising (PCS)		1.71 (0.24)	1.48 (0.24)	3.23 (0.48)	3.76 (0.48)
Fear avoidance (TSK)		-0.34 (0.24)	-0.32 (0.24)	-0.25 (0.48)	0.08 (0.48)
Anxiety (HADS A)		0.96 (0.24)	0.84 (0.24)	0.88 (0.48)	1.85 (0.48)
Depression (HADS D)		1.23 (0.24)	0.76 (0.24)	0.50 (0.48)	0.72 (0.48)
Pain Intensity	Least	1.37 (0.24)	0.80 (0.24)	2.50 (0.48)	0.78 (0.48)
	Average	0.62 (0.24)	0.59 (0.24)	-0.68 (0.48)	-0.40 (0.48)
	Current	1.30 (0.24)	1.02 (0.24)	1.14 (0.48)	1.46 (0.48)

Four-month follow-up scores: Treatment/Medium Risk

Variable		Skewness (SE)		Kurtosis (SE)	
		Four-month Score	**Residualised Score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		1.01 (0.18)	0.41 (0.18)	-0.17 (0.36)	-0.41 (0.36)
General Physical Health (SF-12 PCS)		-0.84 (0.18)	-0.16 (0.18)	0.24 (0.36)	-0.37 (0.36)
Catastrophising (PCS)		1.30 (0.18)	0.88 (0.18)	1.08 (0.36)	0.67 (0.36)
Fear avoidance (TSK)		0.27 (0.18)	-0.05 (0.18)	2.39 (0.36)	1.19 (0.36)
Anxiety (HADS A)		0.70 (0.18)	0.19 (0.18)	-0.08 (0.36)	0.10 (0.36)
Depression (HADS D)		1.32 (0.18)	0.71 (0.18)	1.84 (0.36)	1.88 (0.36)
Pain Intensity		Least	0.97 (0.18)	0.73 (0.18)	0.63 (0.36)
Pain Intensity	Average	0.43 (0.18)	0.32 (0.18)	-0.87 (0.36)	0.52 (0.36)
	Current	0.83 (0.18)	0.63 (0.18)	-0.40 (0.36)	-0.94 (0.36)

Four-month follow-up scores: Treatment/High Risk

Variable		Skewness (SE)		Kurtosis (SE)	
		Original Score	**Residualised Score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		0.59 (0.25)	0.27 (0.25)	-1.03 (0.50)	-0.87 (0.50)
General Physical Health (SF-12 PCS)		-0.08 (0.25)	0.40 (0.25)	-0.56 (0.50)	0.27 (0.50)
Catastrophising (PCS)		0.88 (0.25)	0.56 (0.25)	-0.23 (0.50)	0.30 (0.50)
Fear avoidance (TSK)		-0.04 (0.25)	-0.31 (0.25)	0.99 (0.50)	1.46 (0.50)
Anxiety (HADS A)		0.51 (0.25)	0.35 (0.25)	-0.28 (0.50)	0.43 (0.50)
Depression (HADS D)		0.93 (0.25)	0.59 (0.25)	0.18 (0.50)	0.85 (0.50)
Pain Intensity	Least	0.96 (0.25)	1.00 (0.25)	0.44 (0.50)	0.60 (0.50)
	Average	0.67 (0.25)	0.72 (0.25)	-0.48 (0.50)	0.14 (0.50)
	Current	0.83 (0.25)	0.50 (0.25)	-0.38 (0.50)	-0.38 (0.50)

Four-month follow-up scores: Control/Low Risk

Variable		Skewness (SE)		Kurtosis (SE)	
		Original Score	**Residualised Score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		2.48 (0.34)	1.78 (0.34)	8.94 (0.66)	6.08 (0.66)
General Physical Health (SF-12 PCS)		-0.96 (0.34)	-1.46 (0.34)	1.06 (0.66)	3.39 (0.66)
Catastrophising (PCS)		1.20 (0.34)	1.33 (0.34)	1.19 (0.66)	1.90 (0.66)
Fear avoidance (TSK)		-0.19 (0.34)	-0.24 (0.34)	1.19 (0.66)	0.70 (0.66)
Anxiety (HADS A)		0.50 (0.34)	-0.24 (0.34)	-0.69 (0.66)	0.35 (0.66)
Depression (HADS D)		1.20 (0.34)	1.07 (0.34)	1.02 (0.66)	3.21 (0.66)
Pain Intensity	Least	0.54 (0.34)	0.59 (0.34)	-1.15 (0.66)	-0.78 (0.66)
	Average	0.20 (0.34)	0.28 (0.34)	-1.11 (0.66)	-0.70 (0.66)
	Current	0.88 (0.34)	0.94 (0.34)	0.18 (0.66)	0.55 (0.66)

Four-month follow-up scores: Control/Medium Risk

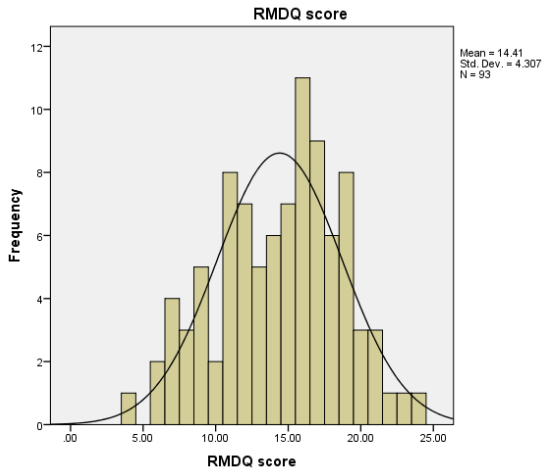
Variable		Skewness (SE)		Kurtosis (SE)	
		Original Score	**Residualised Score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		0.79 (0.26)	0.50 (0.26)	-0.76 (0.52)	-0.69 (0.52)
General Physical Health (SF-12 PCS)		-0.21 (0.26)	0.19 (0.26)	-0.82 (0.52)	-0.14 (0.52)
Catastrophising (PCS)		1.21 (0.26)	0.94 (0.26)	1.77 (0.52)	1.82 (0.52)
Fear avoidance (TSK)		-0.44 (0.26)	-0.16 (0.26)	0.23 (0.52)	0.18 (0.52)
Anxiety (HADS A)		0.56 (0.26)	-0.11 (0.26)	-0.21 (0.52)	0.53 (0.52)
Depression (HADS D)		1.45 (0.26)	0.95 (0.26)	2.45 (0.52)	1.43 (0.52)
Pain Intensity	Least	0.75 (0.26)	0.65 (0.26)	-0.26 (0.52)	0.25 (0.52)
	Average	0.39 (0.26)	0.35 (0.26)	-0.72 (0.52)	-0.55 (0.52)
	Current	0.57 (0.26)	0.53 (0.26)	-0.77 (0.52)	-0.69 (0.52)

Four-month follow-up scores: Control/High Risk

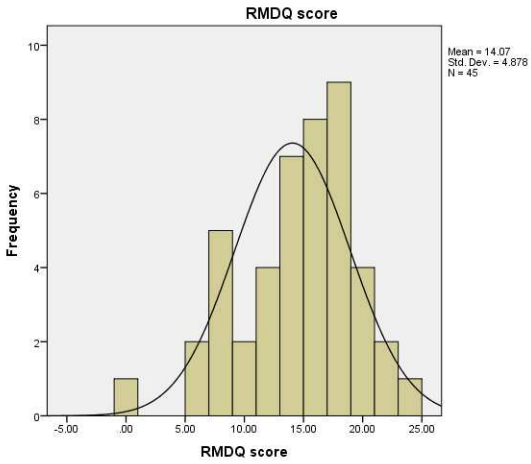
Variable		Skewness (SE)		Kurtosis (SE)	
		Original Score	**Residualised change score**	**Original Score**	**Residualised Score**
Disability (RMDQ)		-0.06 (0.35)	0.04 (0.35)	-1.11 (0.70)	-0.98 (0.70)
General Physical Health (SF-12 PCS)		0.48 (0.35)	0.68 (0.35)	-0.85 (0.70)	-0.08 (0.70)
Catastrophising (PCS)		0.37 (0.35)	0.15 (0.35)	-0.68 (0.70)	-1.08 (0.70)
Fear avoidance (TSK)		0.28 (0.35)	-0.24 (0.35)	0.89 (0.70)	0.10 (0.79)
Anxiety (HADS A)		0.61 (0.35)	-0.13 (0.35)	0.47 (0.70)	0.11 (0.70)
Depression (HADS D)		0.62 (0.35)	-0.04 (0.35)	0.35 (0.70)	-0.86 (0.70)
Pain Intensity	Least	0.56 (0.35)	0.53 (0.35)	-0.55 (0.70)	-0.15 (0.70)
	Average	-0.24 (0.35)	-0.22 (0.35)	-0.97 (0.70)	-0.82 (0.70)
	Current	0.00 (0.35)	0.35 (0.35)	-1.21 (0.70)	-0.34 (0.70)

High risk group: Baseline Scores: Histograms

RMDQ

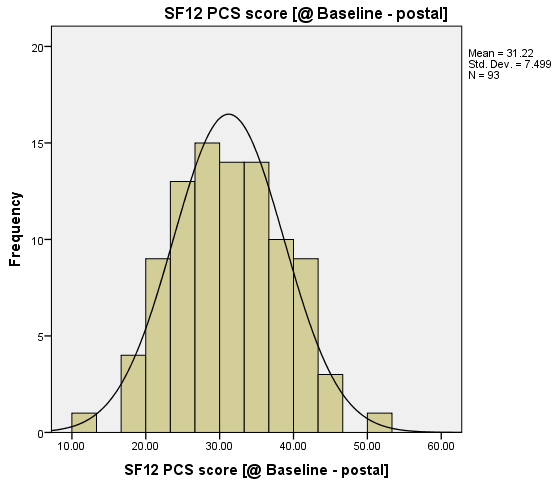


Treatment

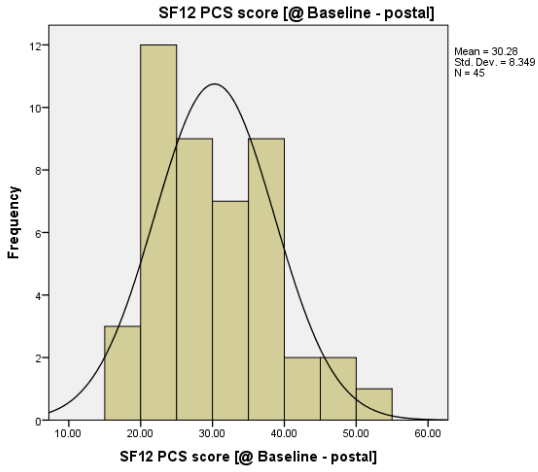


Control

SF-12 PCS

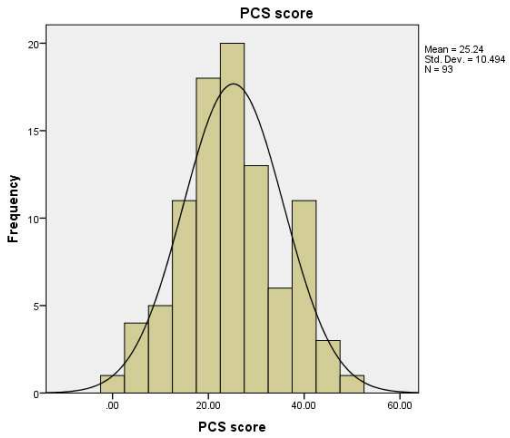


Treatment

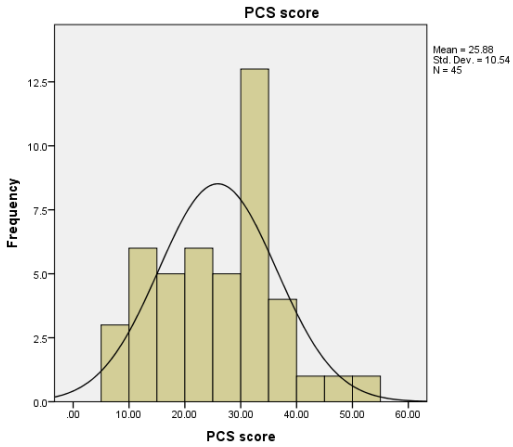


Control

PCS

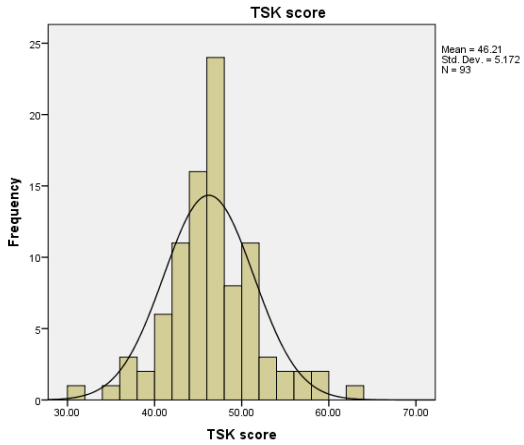


Treatment

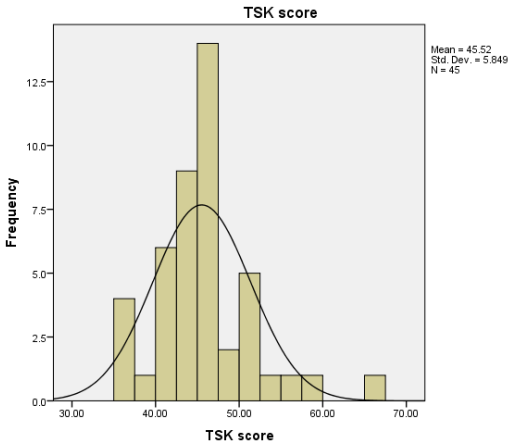


Control

TSK

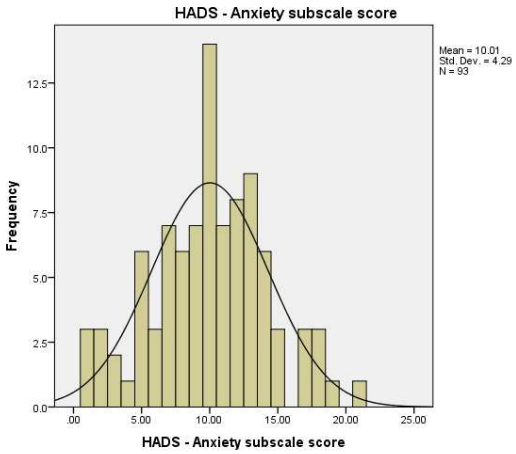


Treatment

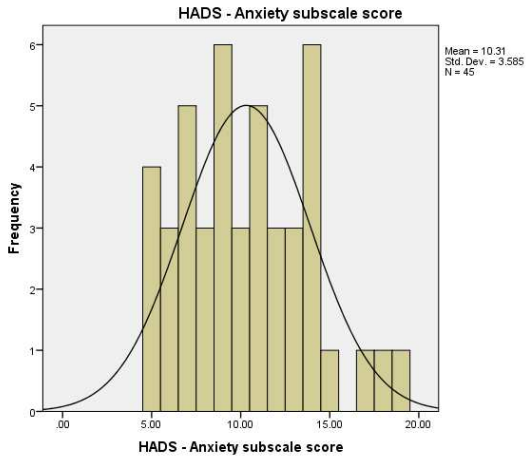


Control

HADS-A

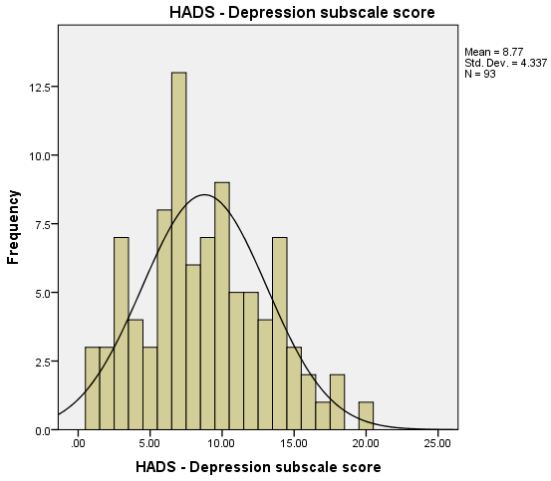


Treatment

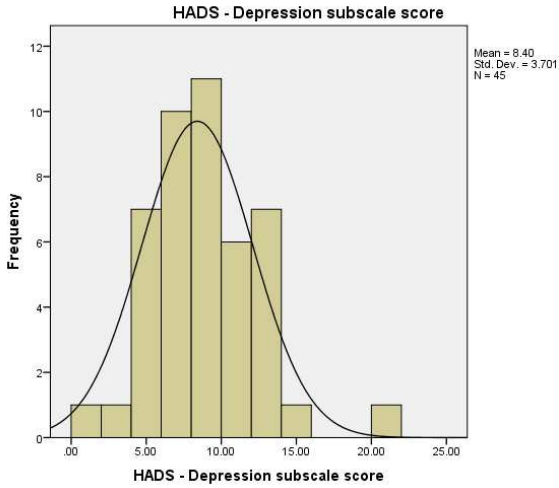


Control

HADS-D

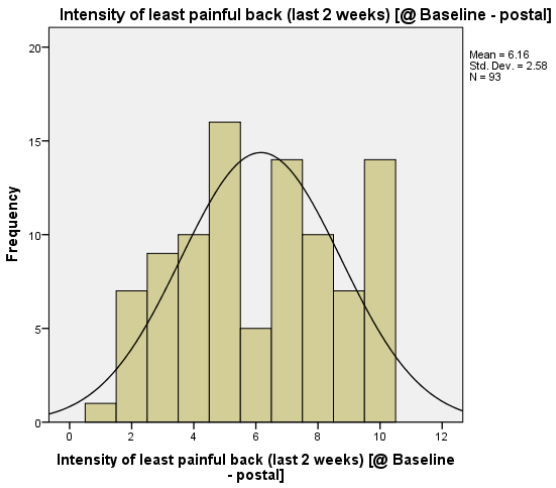


Treatment

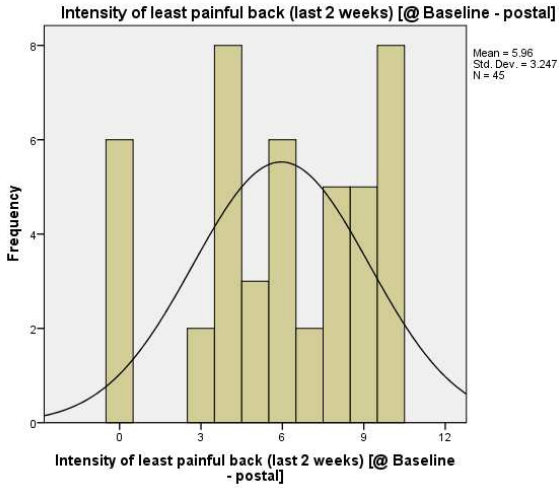


Control

Least Pain

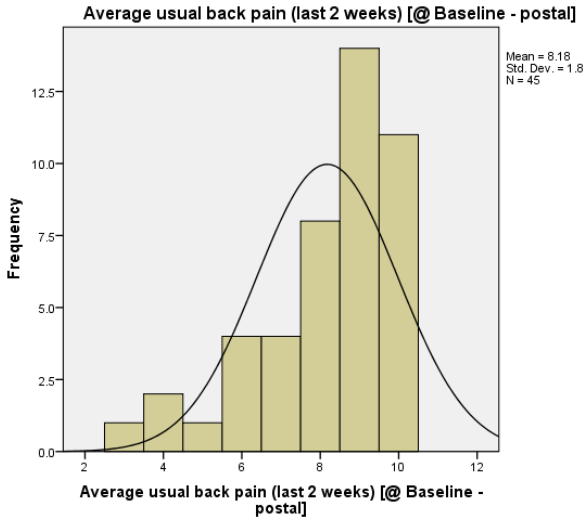


Treatment

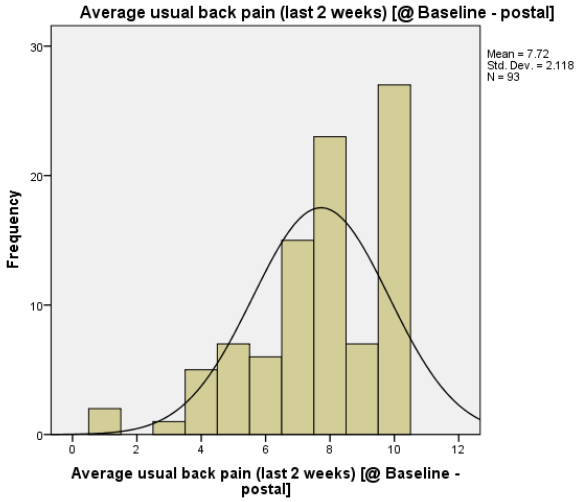


Control

Average Pain

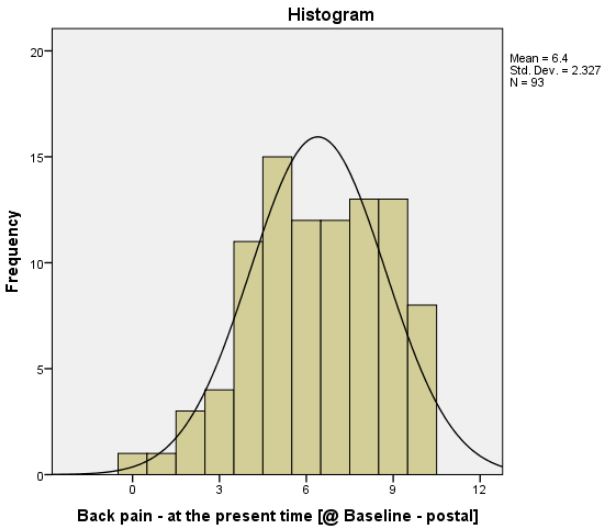


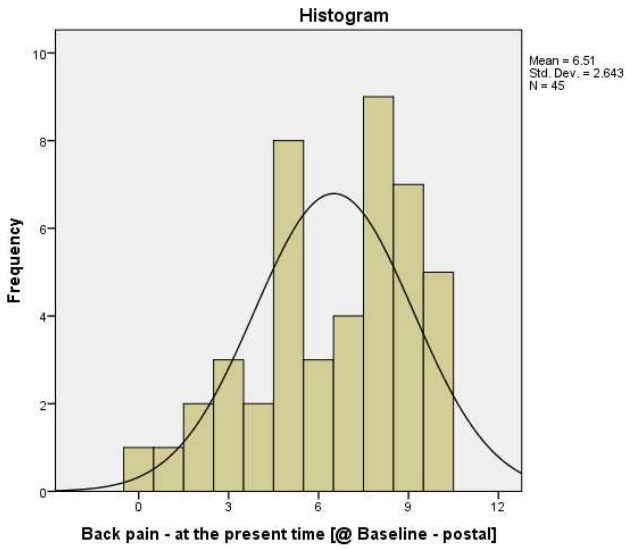
Treatment



Control

Current Pain

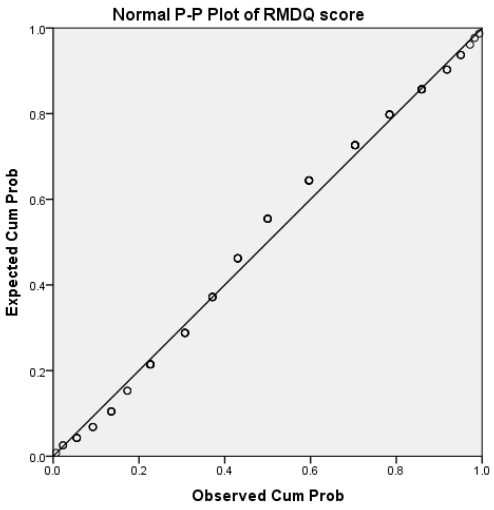


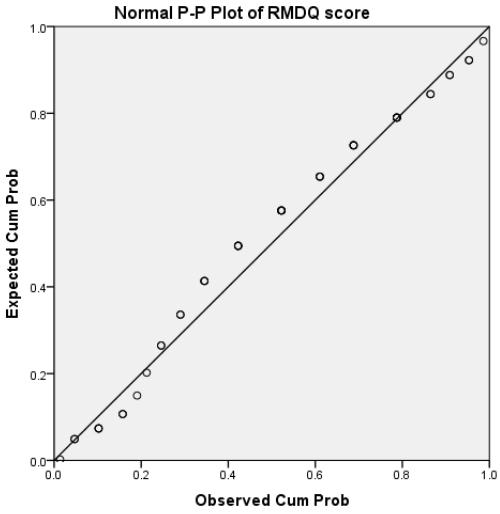


Control

High Risk Group: Baseline Scores: P-Plots

RMDQ

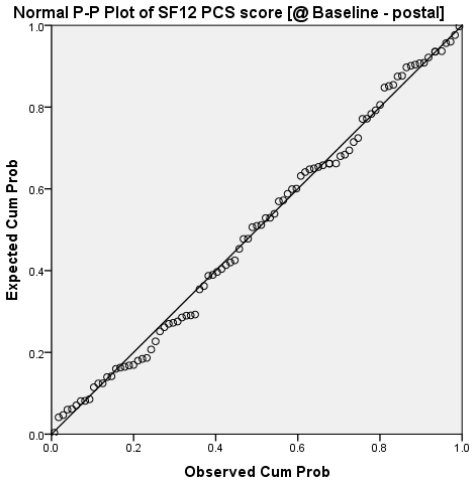




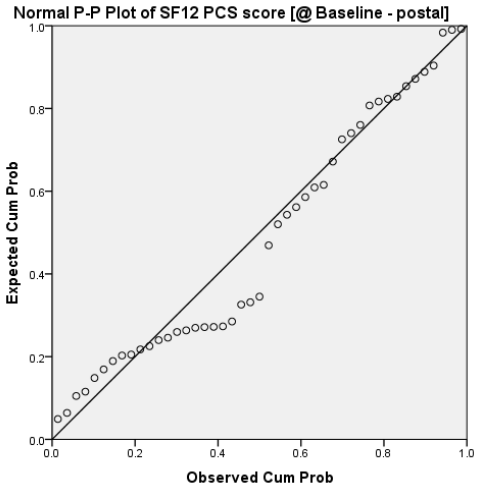
Treatment

Control

SF-12

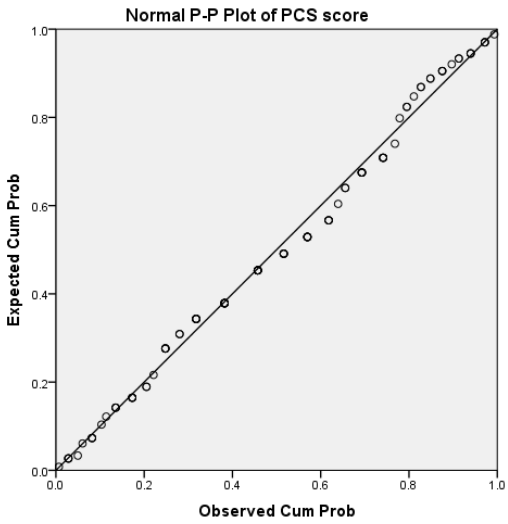


Treatment

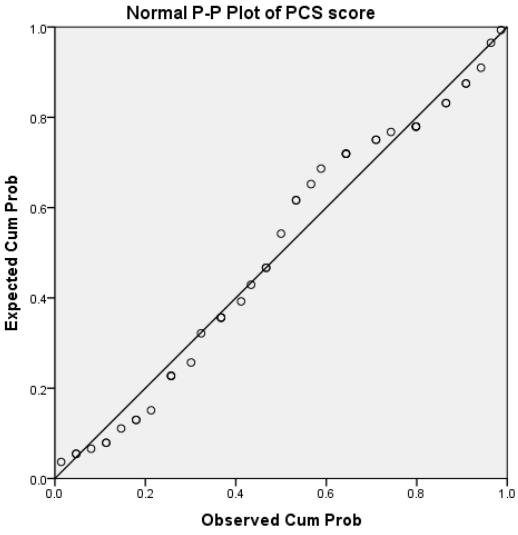


Control

PCS

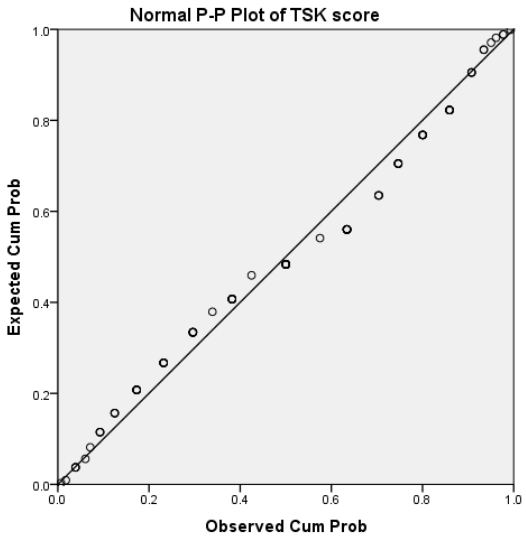


Treatment

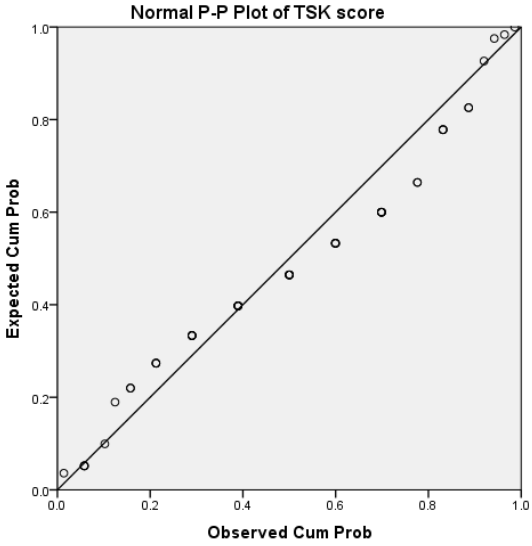


Control

TSK

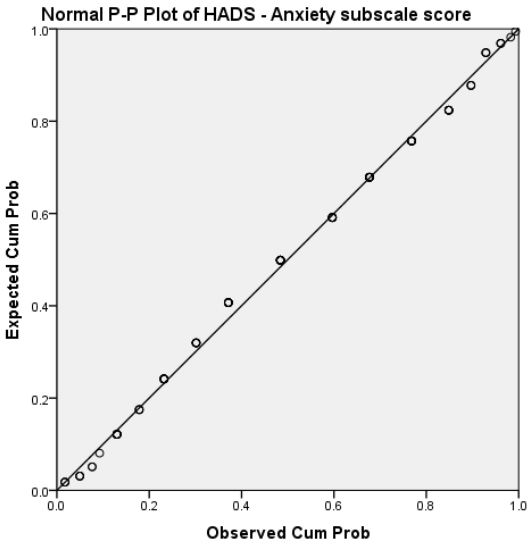


Treatment

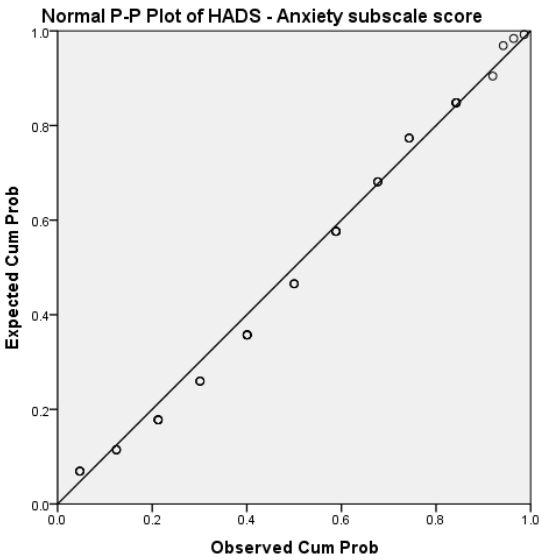


Control

HADS-A

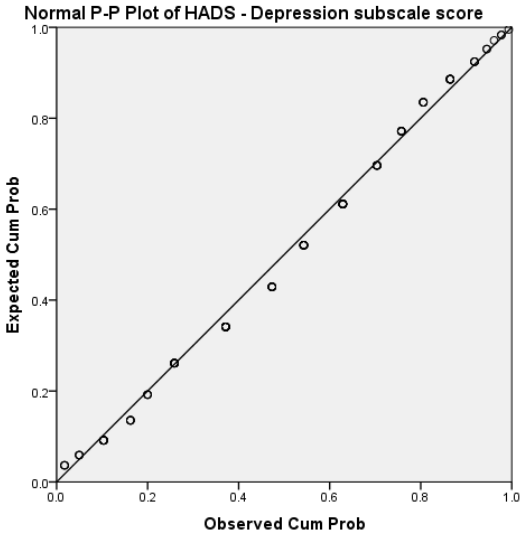


Treatment

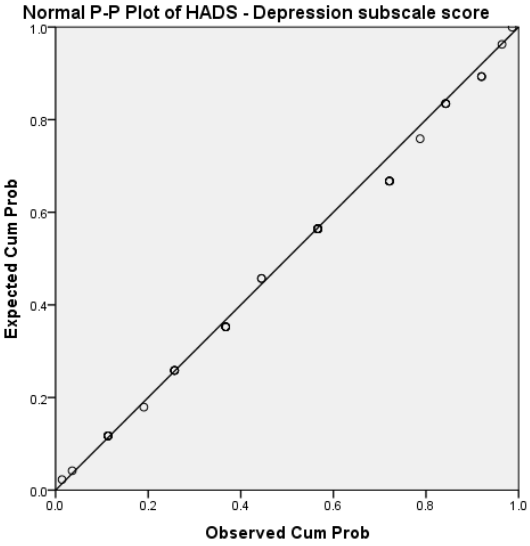


Control

HADS-D

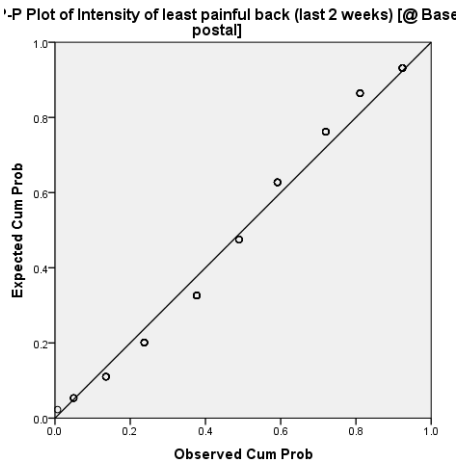


Treatment

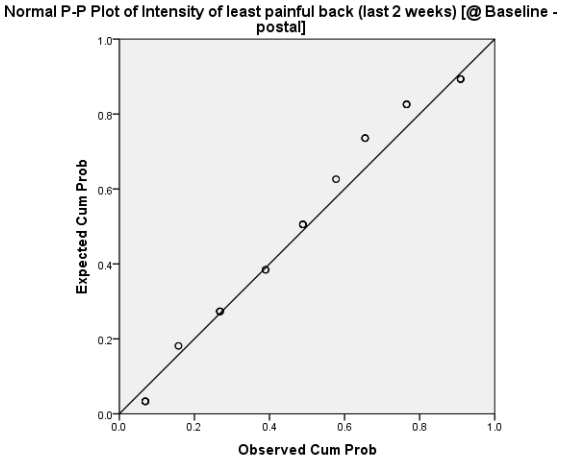


Control

Least Pain

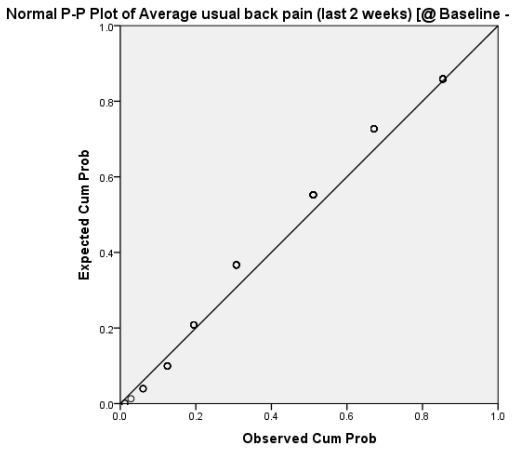


Treatment

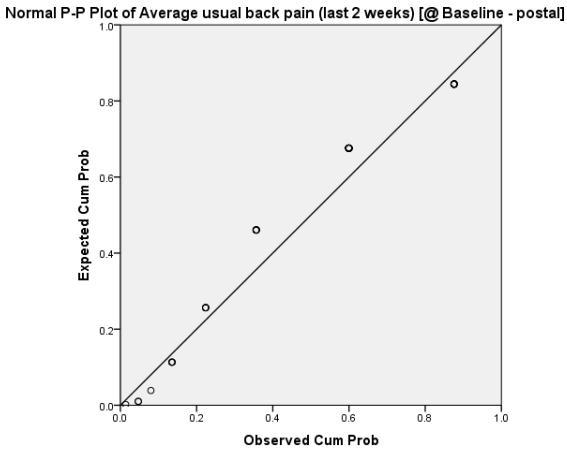


Control

Average Pain

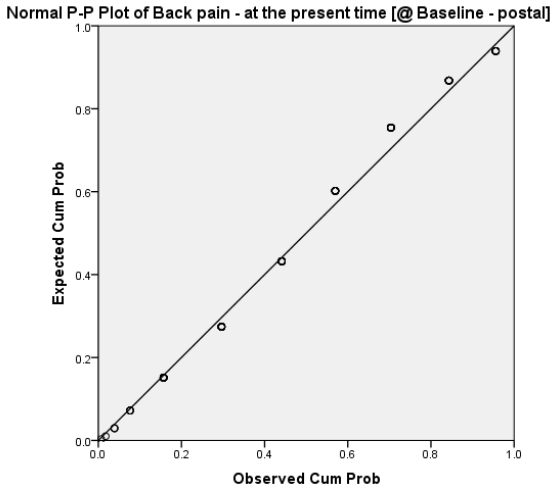


Treatment

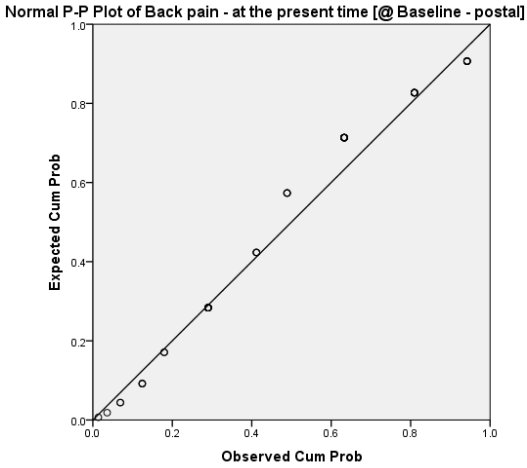


Control

Current Pain



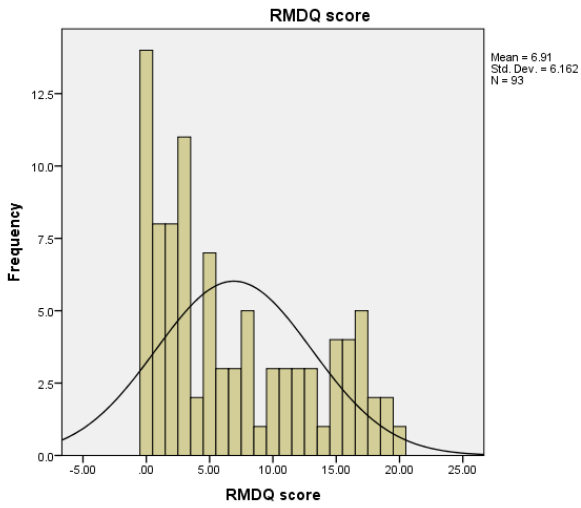
Treatment



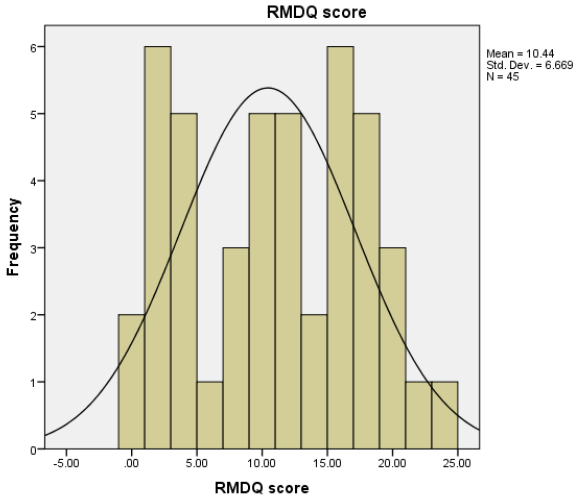
Control

High Risk: Four-month Follow-up Scores: Histograms

RMDQ

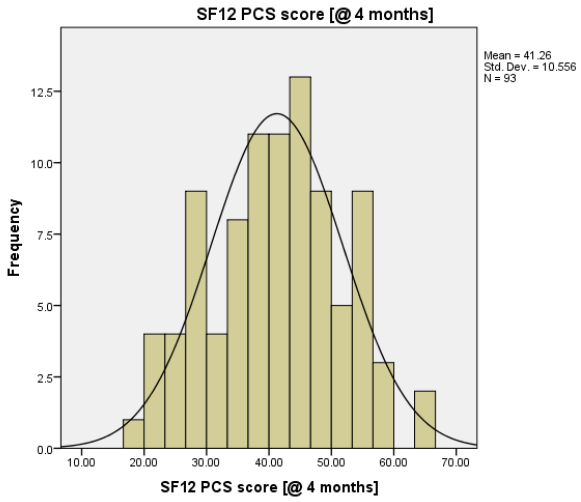


Treatment

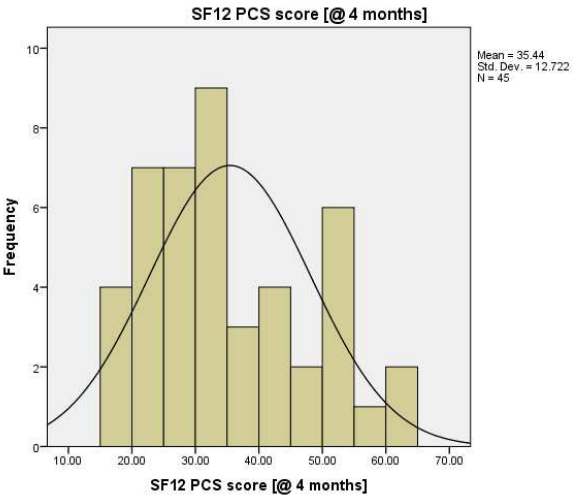


Control

SF-12 PCS

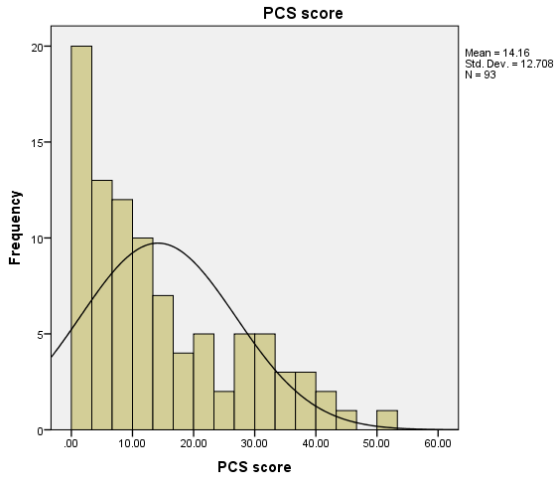


Treatment

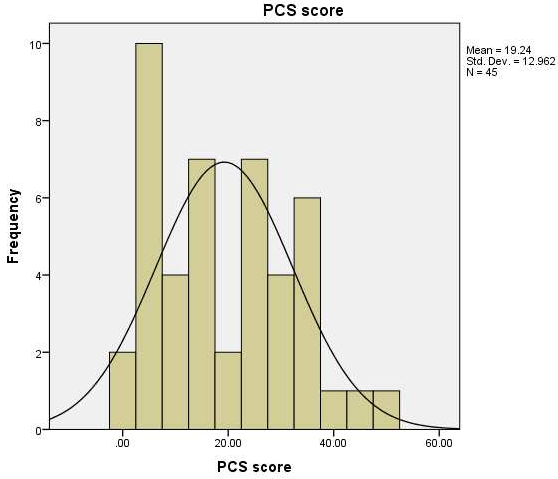


Control

PCS

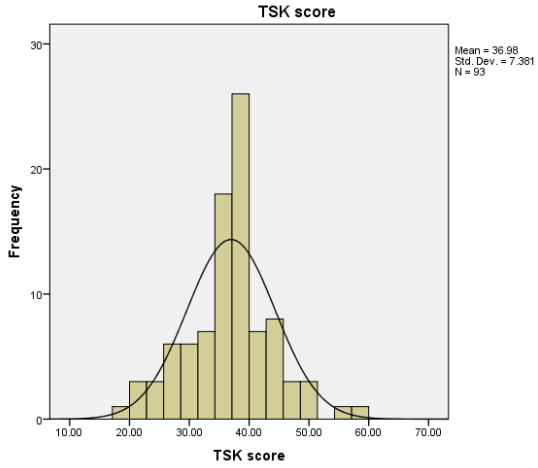


Treatment

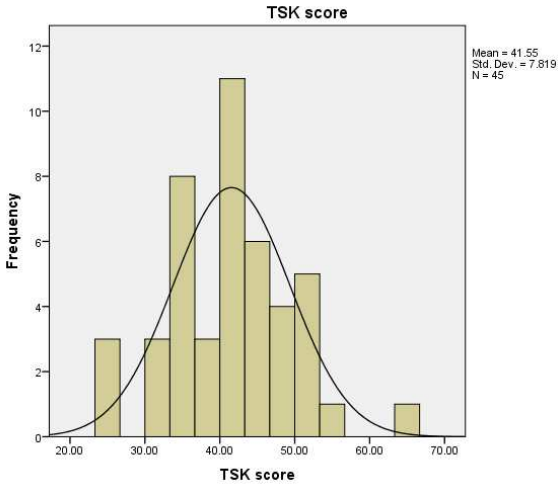


Control

TSK

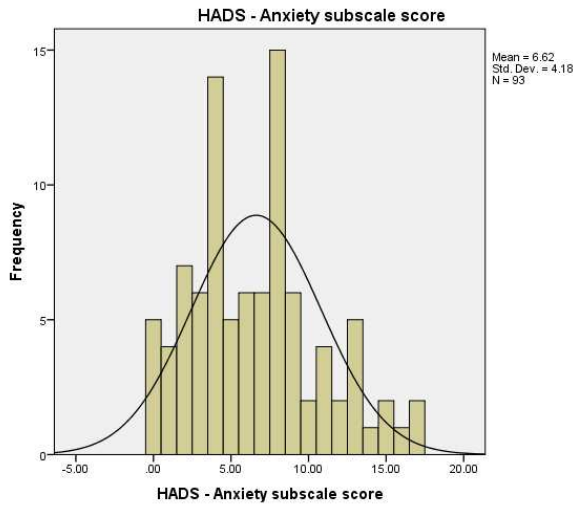


Treatment

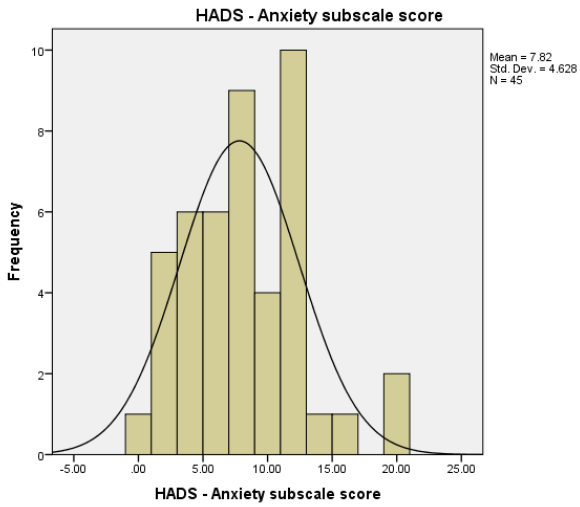


Control

HADS-A

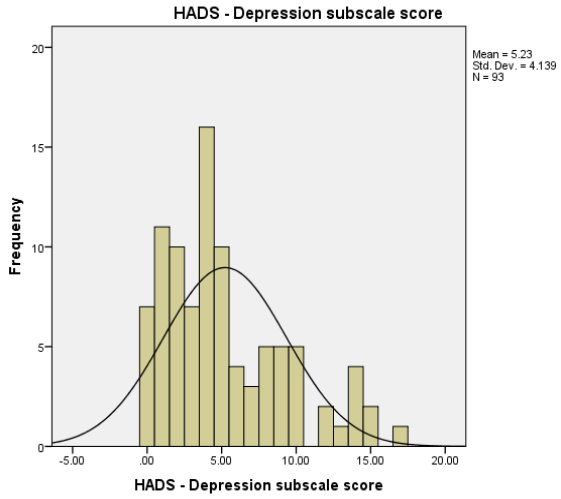


Treatment

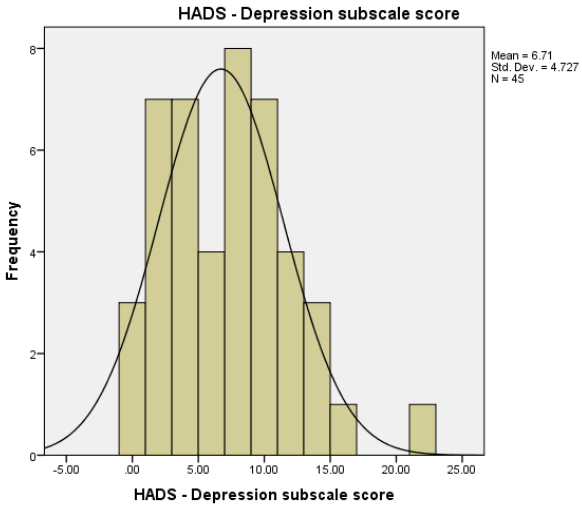


Control

HADS-D

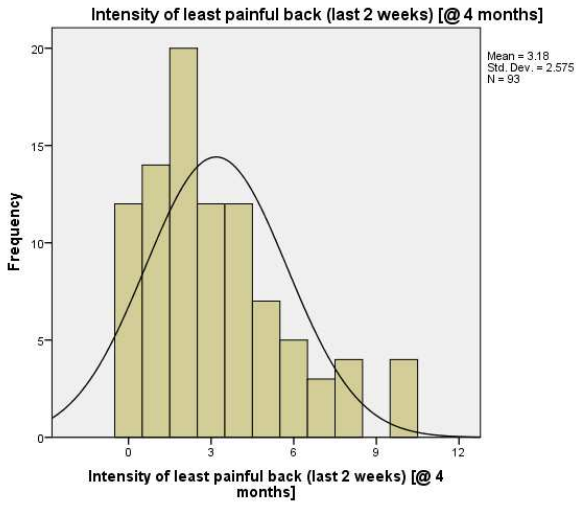


Treatment

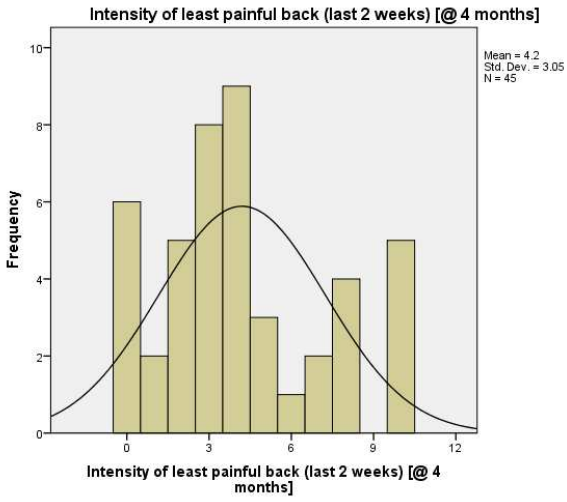


Control

Least Pain

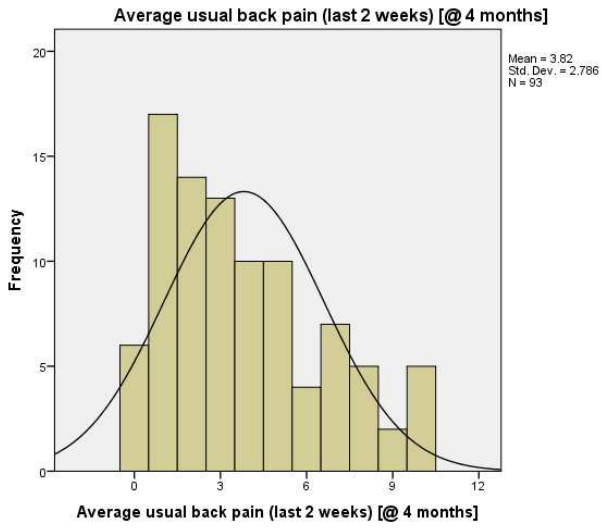


Treatment

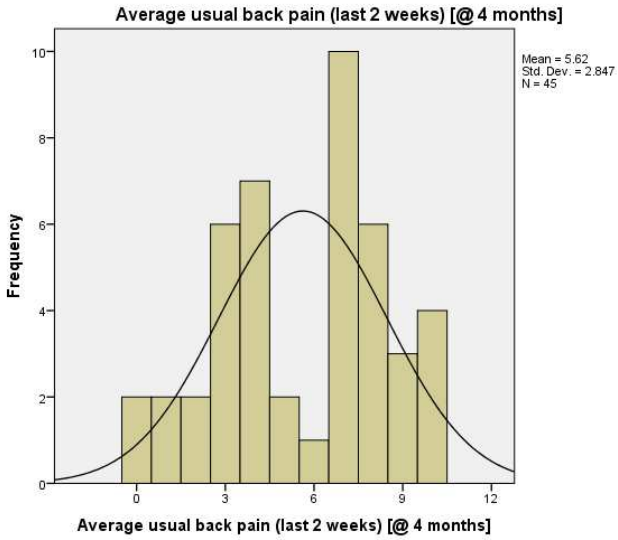


Control

Average Pain

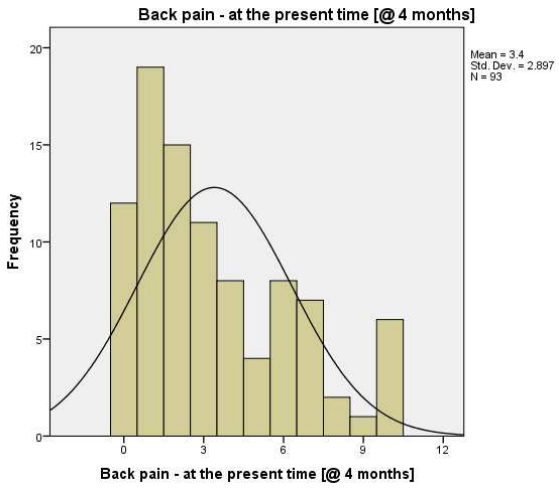


Treatment

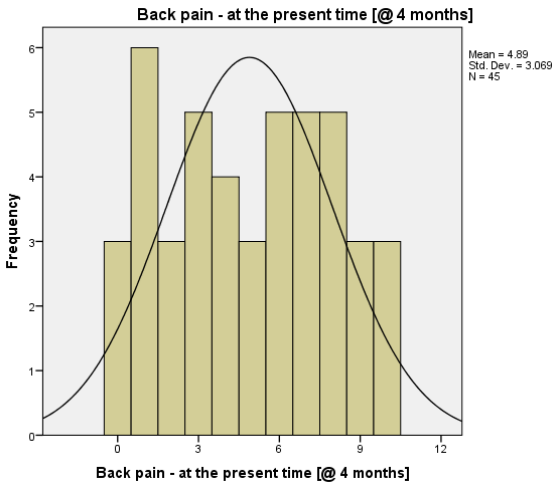


Control

Current Pain



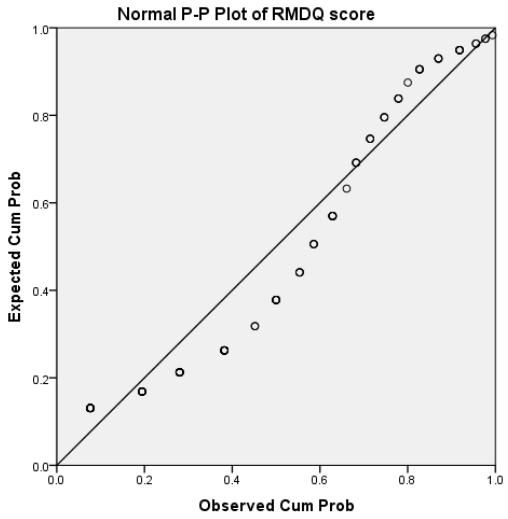
Treatment



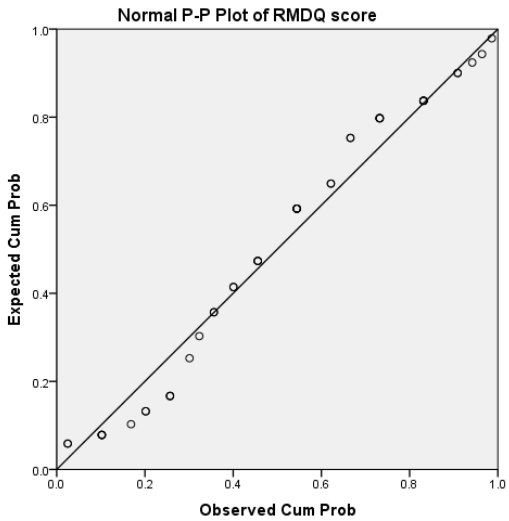
Control

High Risk: 4m Follow-up scores: P-Plots

RMDQ

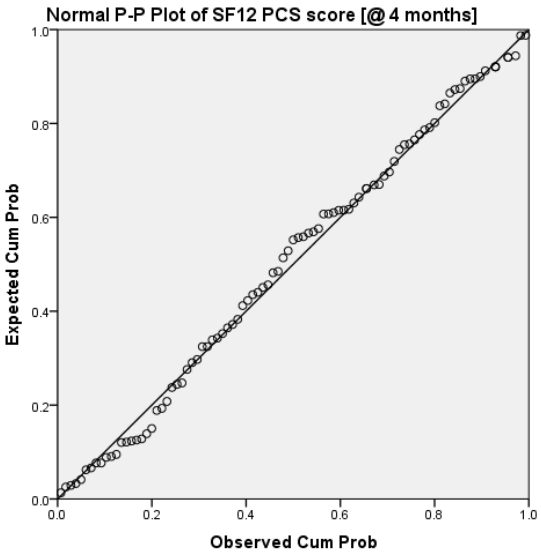


Treatment

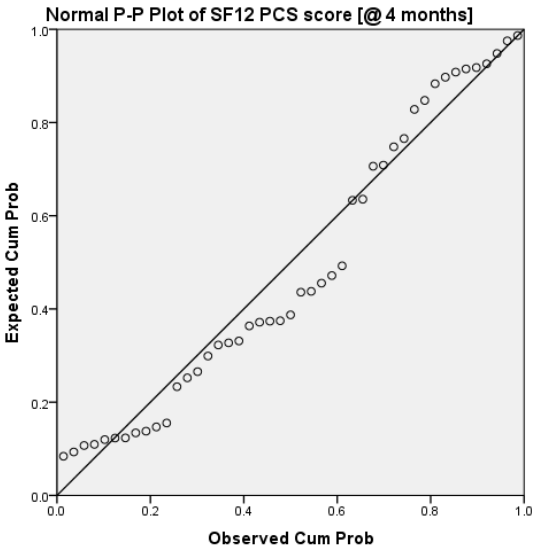


Control

SF-12 PCS

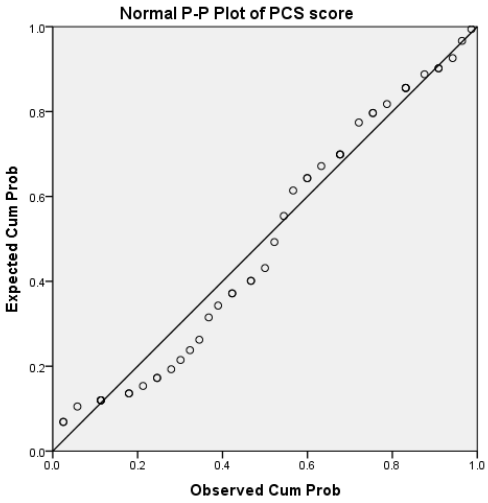


Treatment

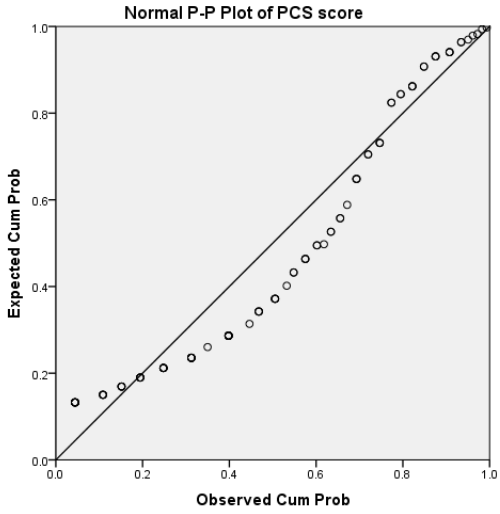


Control

PCS

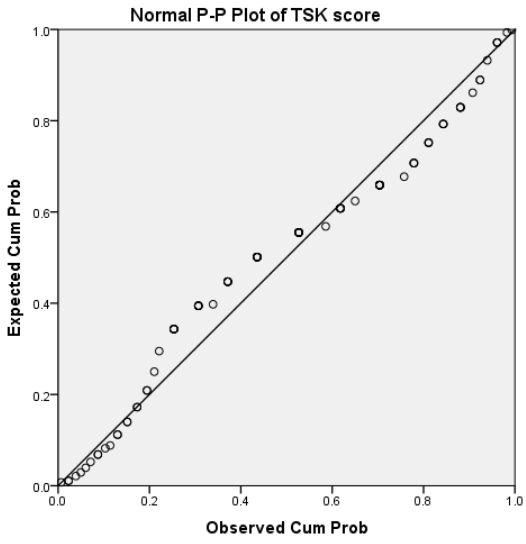


Treatment

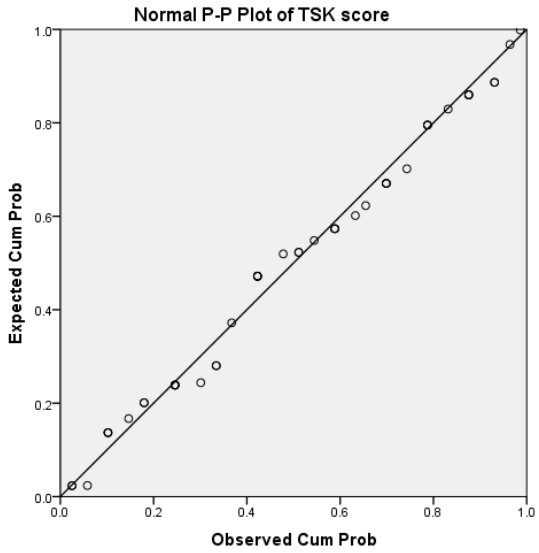


Control

TSK

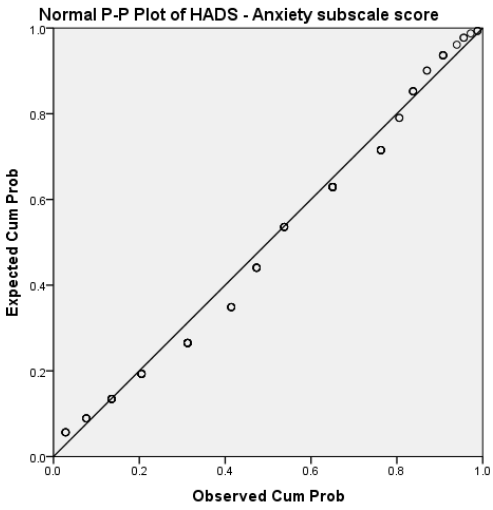


Treatment

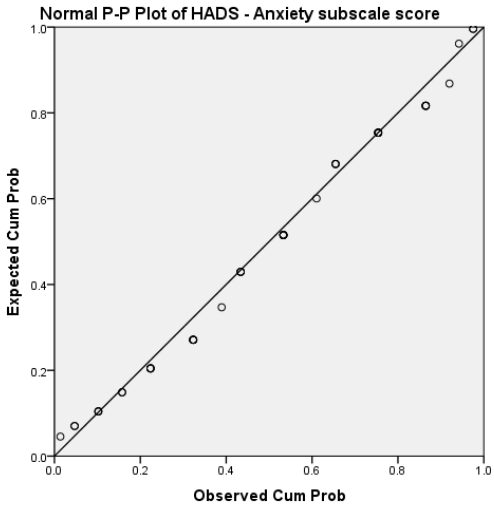


Control

HADS-A

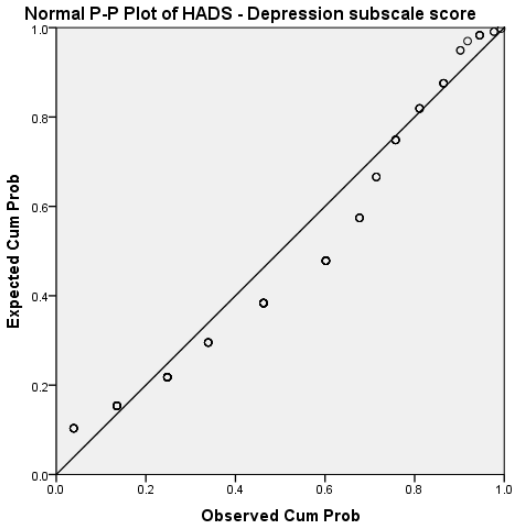


Treatment

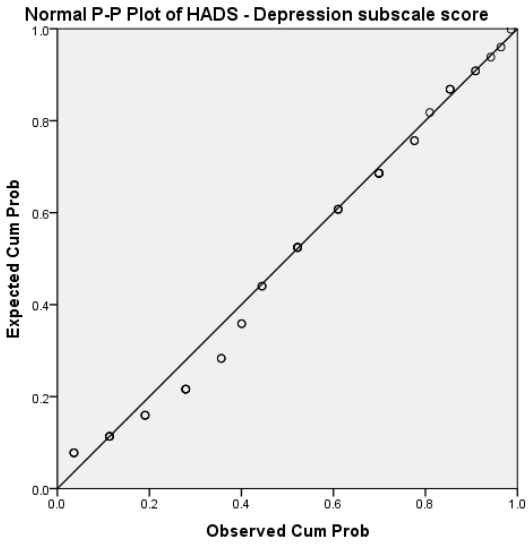


Control

HADS-D

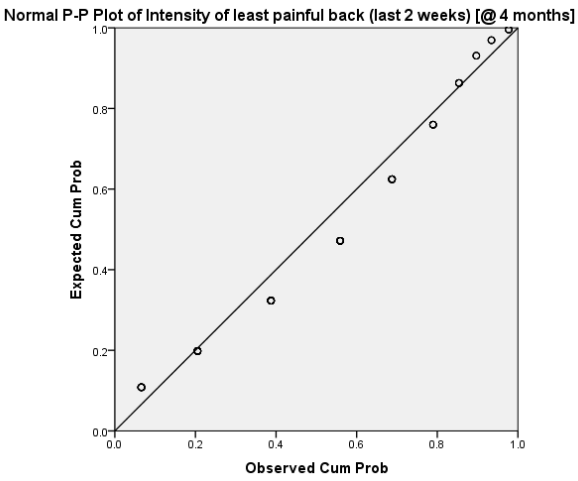


Treatment

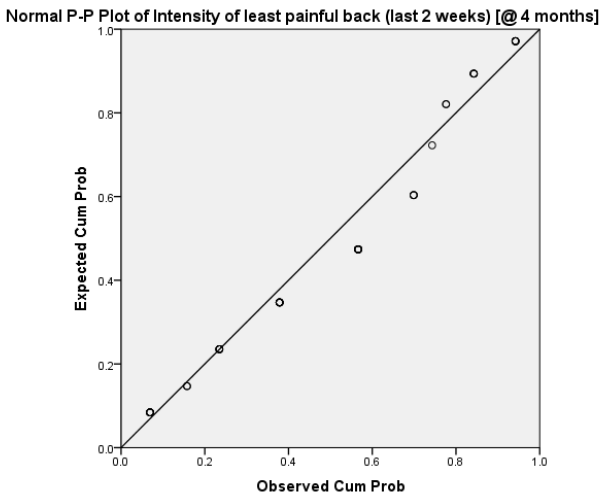


Control

Least Pain

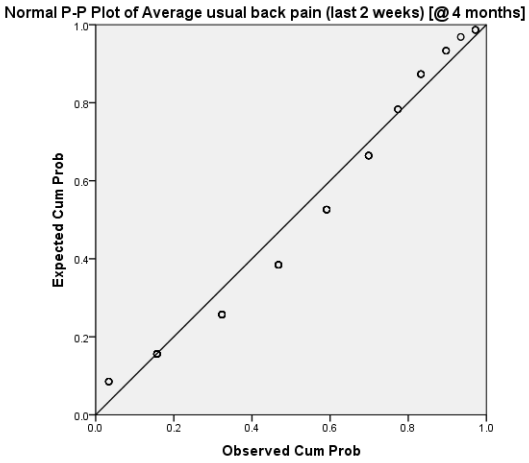


Treatment

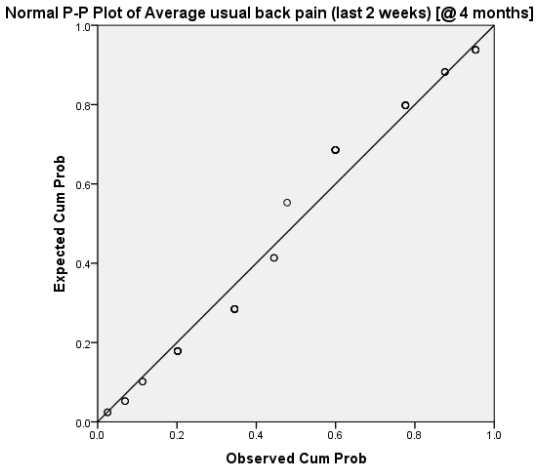


Control

Average Pain

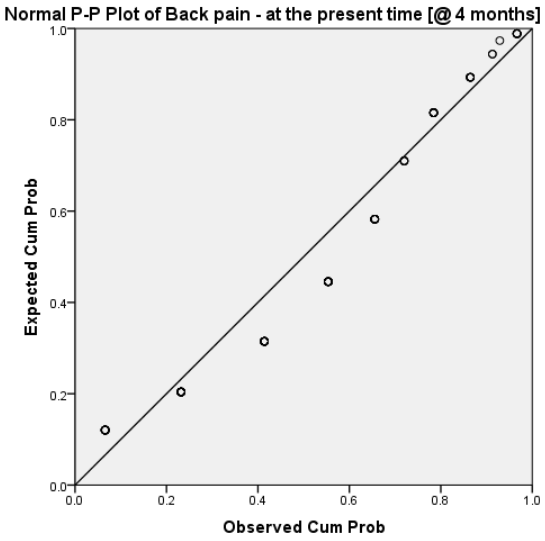


Treatment

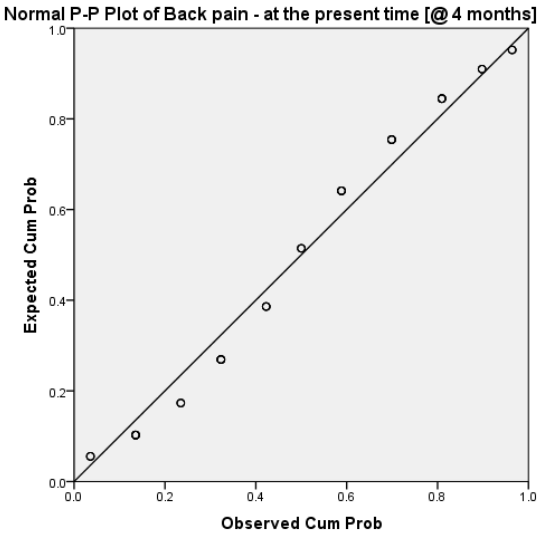


Control

Current Pain



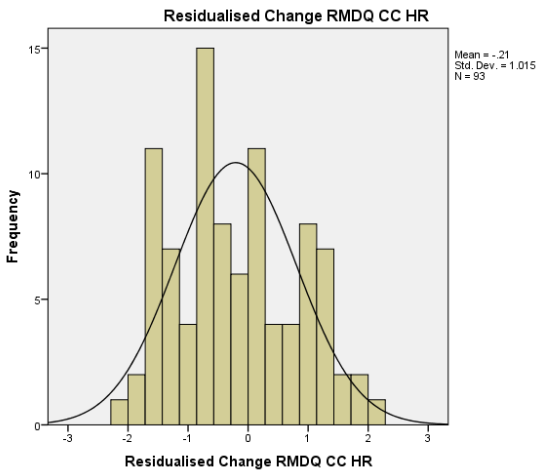
Treatment



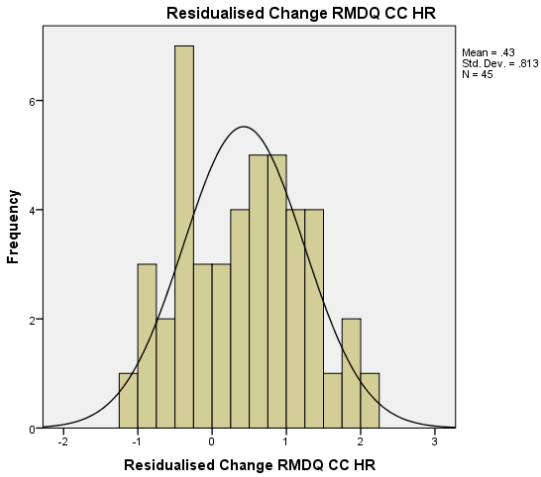
Control

High Risk: Residualised Change Scores: Histograms

RMDQ

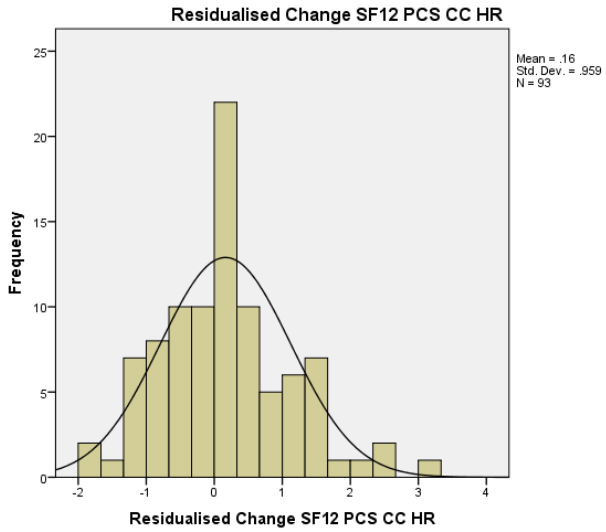


Treatment

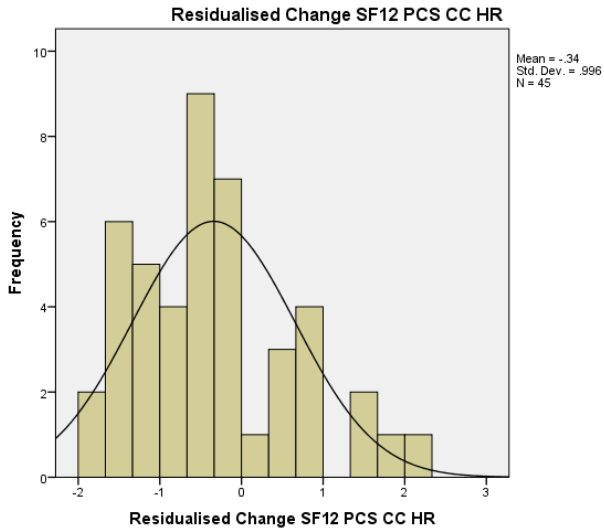


Control

SF-12 PCS

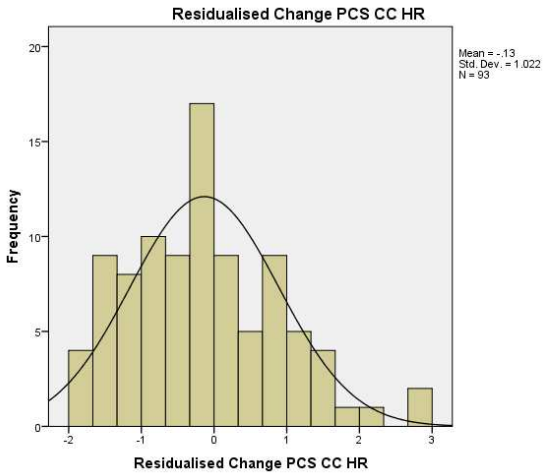


Treatment

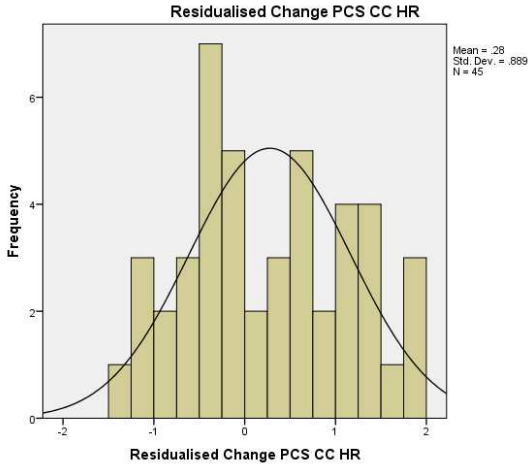


Control

PCS

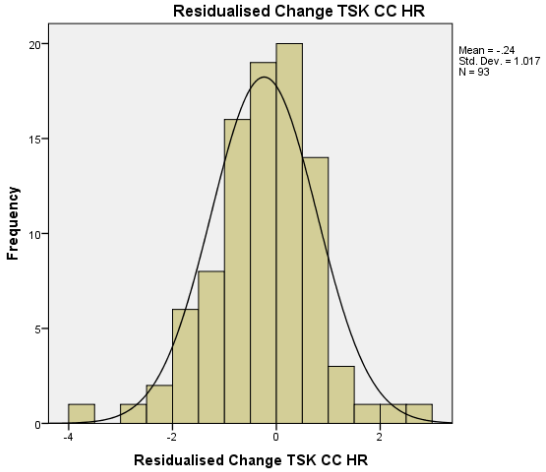


Treatment

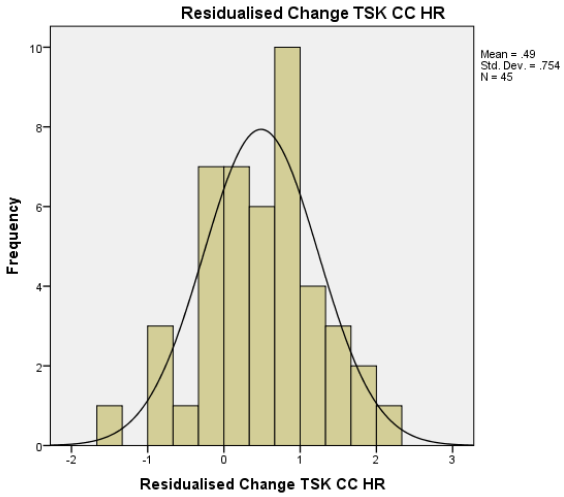


Control

TSK

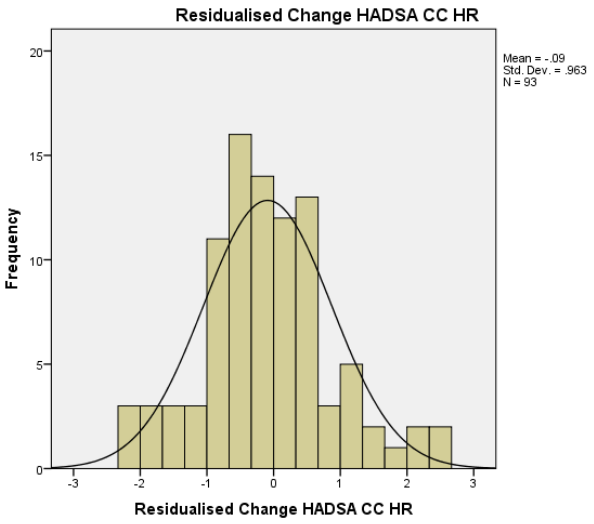


Treatment

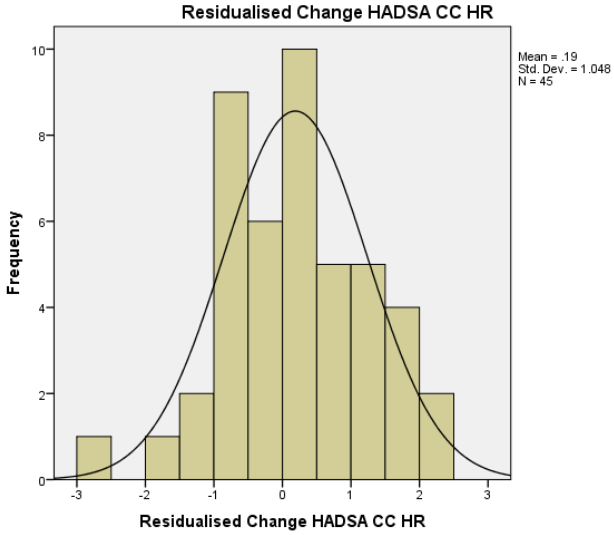


Control

HADS-A

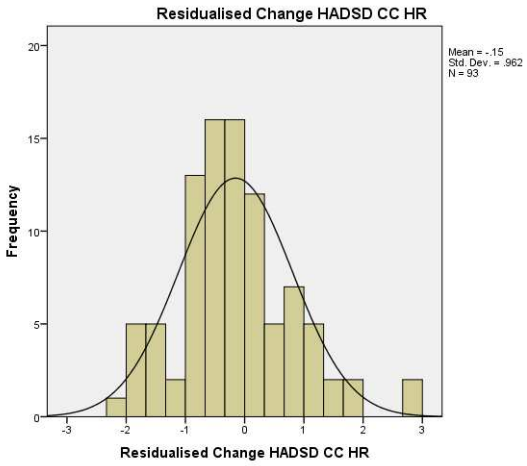


Treatment

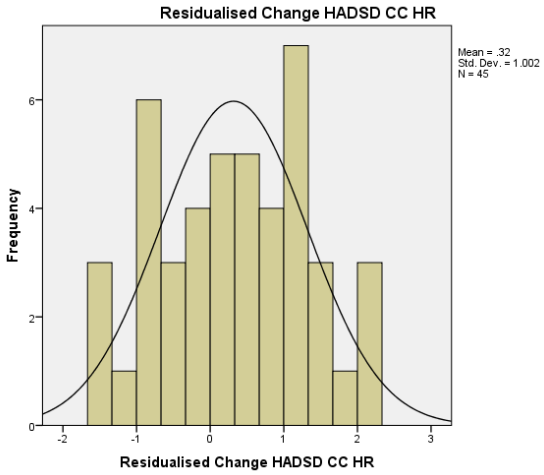


Control

HADS-D

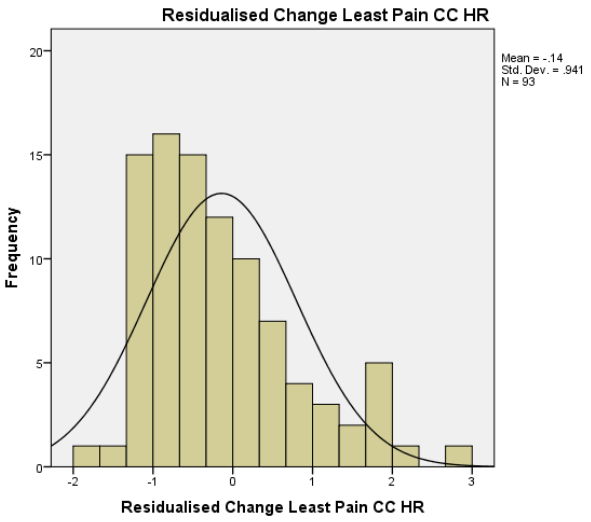


Treatment

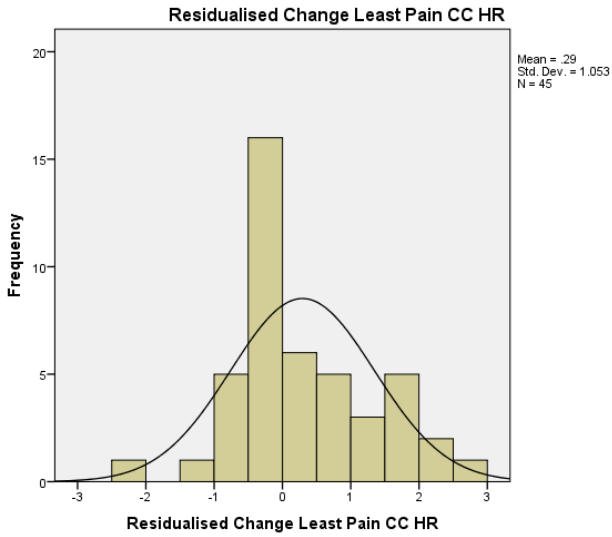


Control

Least Pain

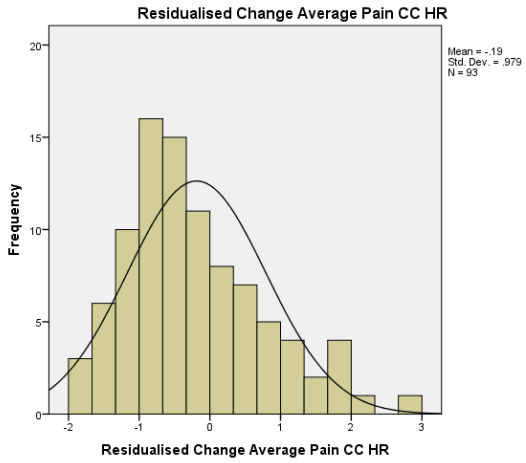


Treatment

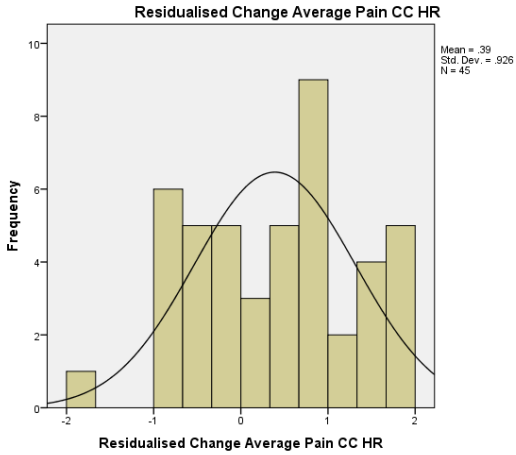


Control

Average Pain

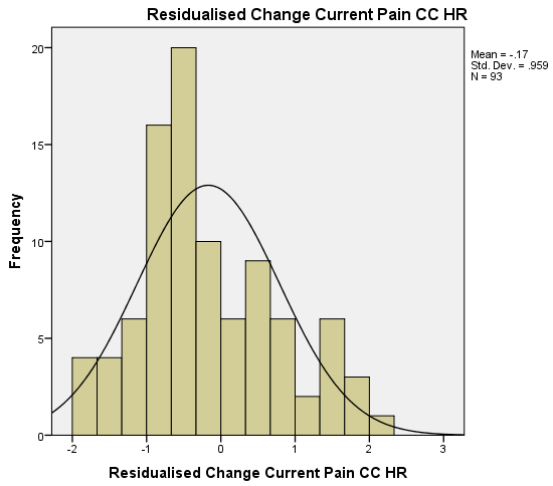


Treatment

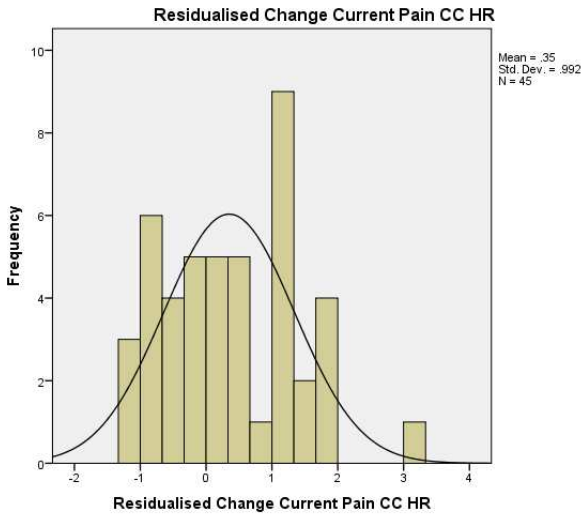


Control

Current Pain



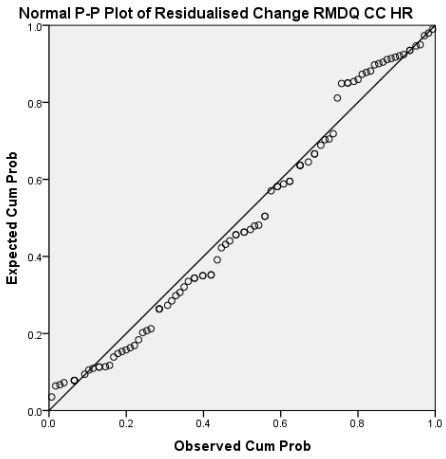
Treatment



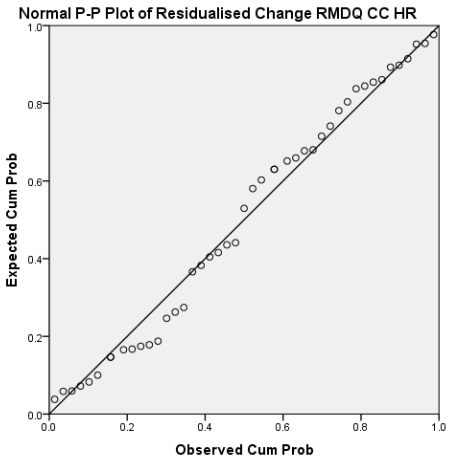
Control

High Risk Group: Residualised Change Scores: P-Plots

RMDQ

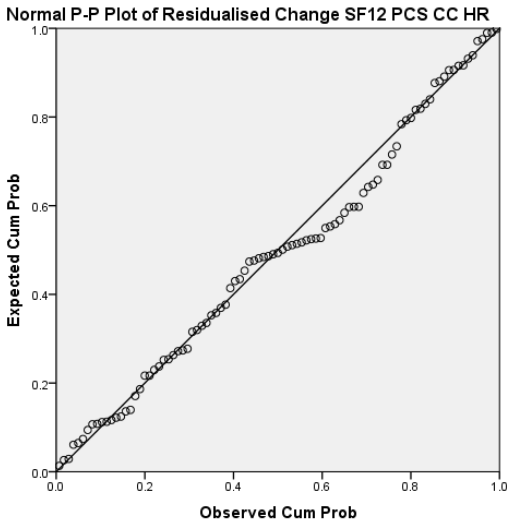


Treatment

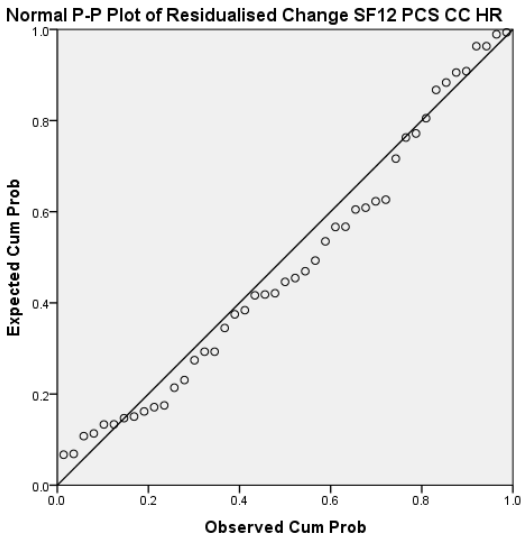


Control

SF-12 PCS

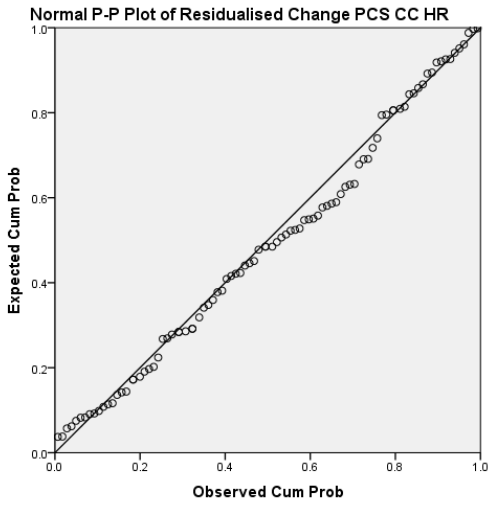


Treatment

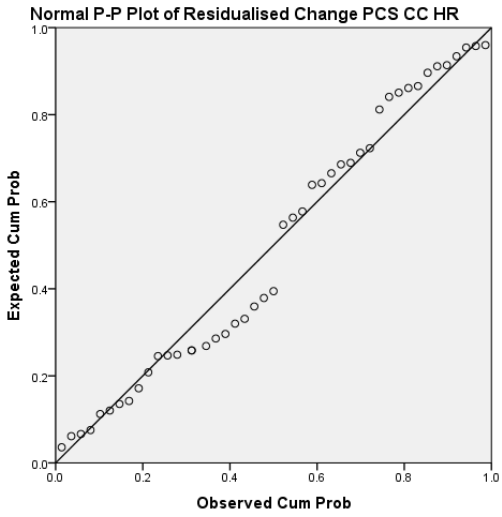


Control

PCS

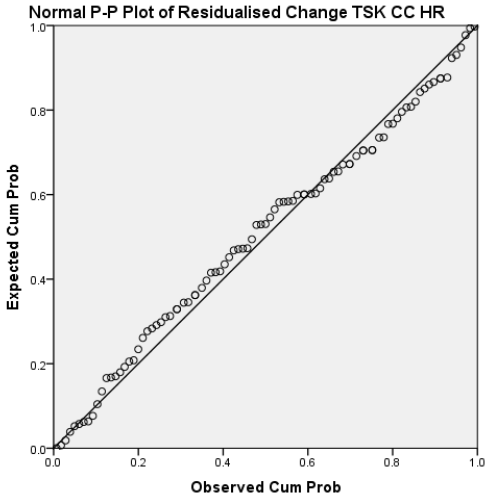


Treatment

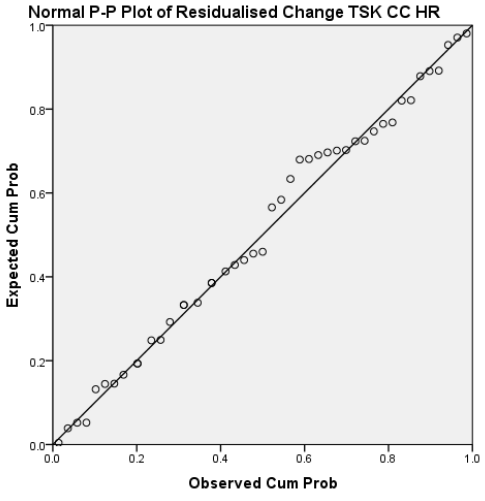


Control

TSK

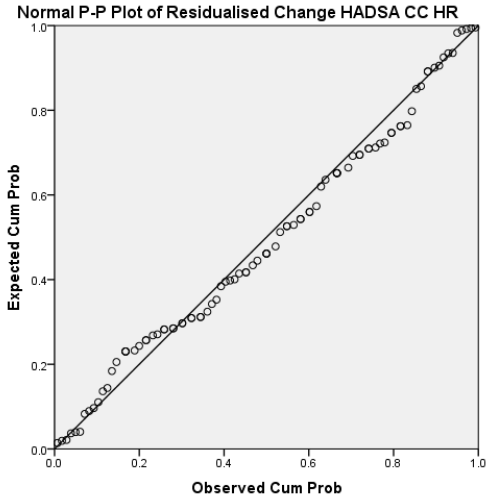


Treatment

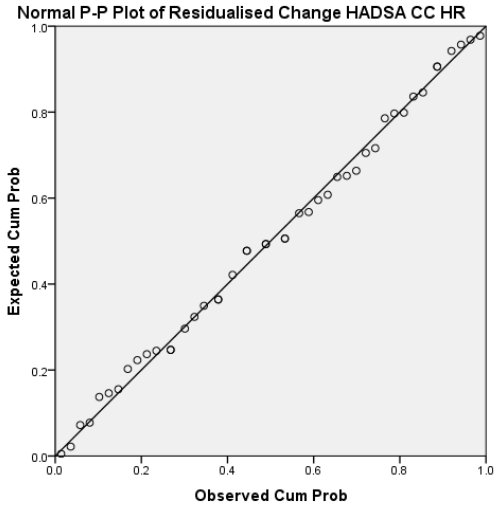


Control

HADS-A

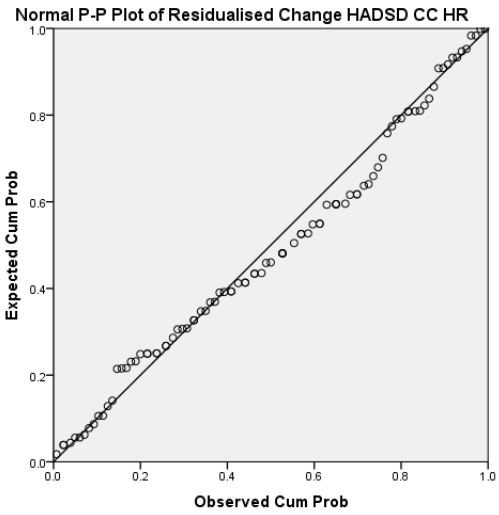


Treatment

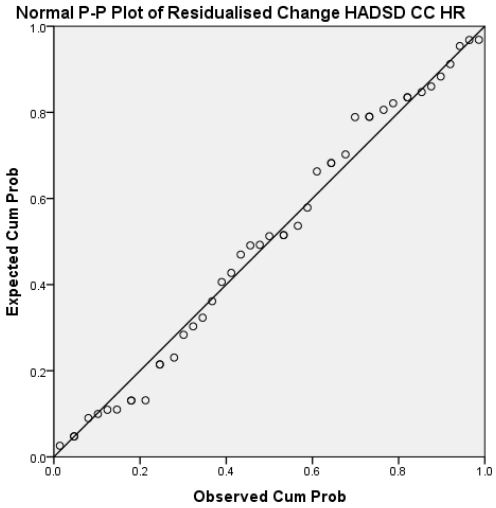


Control

HADS-D

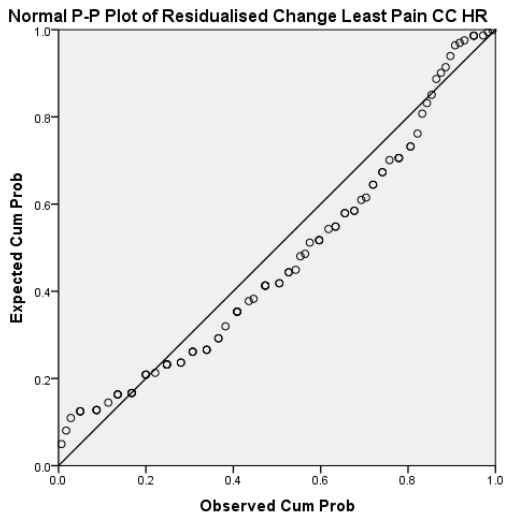


Treatment

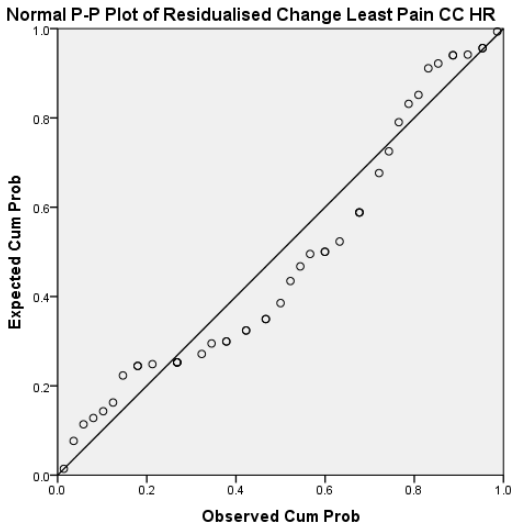


Control

Least Pain

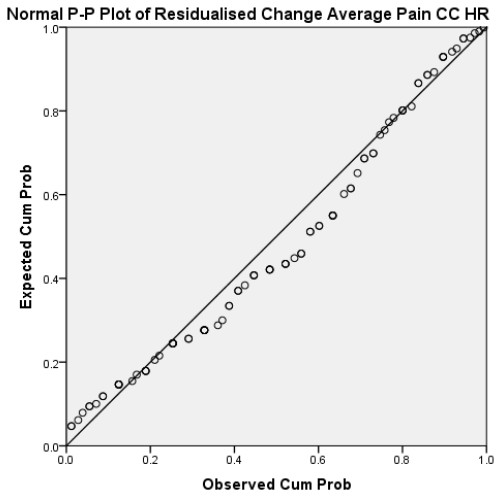


Treatment

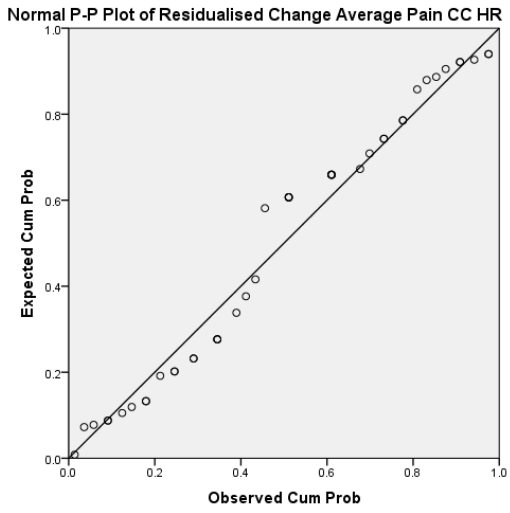


Control

Average Pain

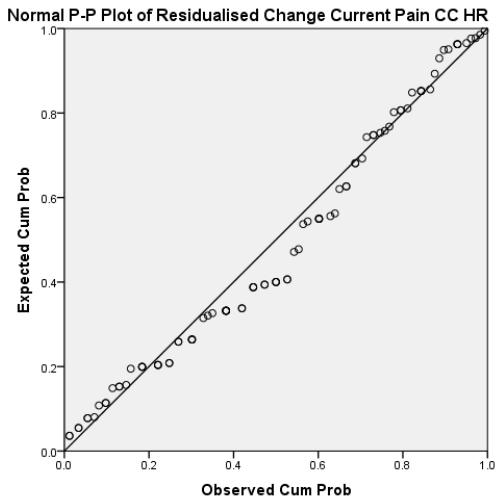


Treatment

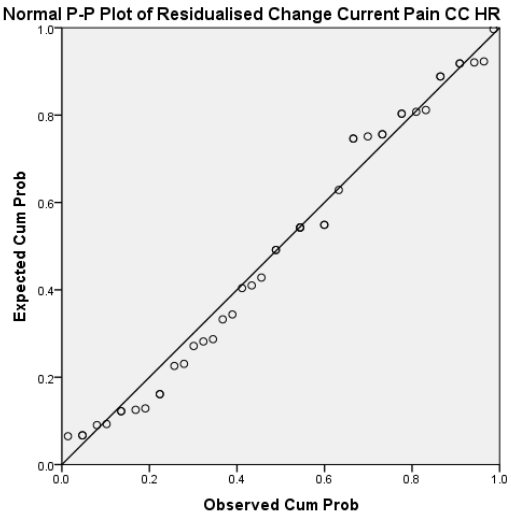


Control

Current Pain



Treatment



Control

Appendix 6.7 Analysis of Potential Non-Response Bias in STarT Back participants (baseline characteristics of responders and non-responders to follow-up questionnaires)

All data

Measure	Baseline values (mean (SD)) (*n*=851)	Baseline values (Mean (SD)) at four-month follow-up		
		Resp (*n*=689)	Non-resp (*n*=162)	*Difference*
Age	**49.73 (14.76)**	**51.88 (14.19)**	**40.60 (13.65)**	**11.28**
Sex	500 (47%) f	410 (60%) f	90 (56%) f	
RMDQ	**9.76 (5.67)**	**9.57 (5.63)**	**10.59 (5.78)**	**1.02**
SF-12 PCS	**36.80 (10.47)**	**37.00 (10.49)**	**35.98 (10.39)**	**1.02**
TSK	40.42 (6.20)	40.31 (6.05)	40.88 (6.79)	0.57
PCS	**16.14 (10.86)**	**15.79 (10.70)**	**17.65 (11.42)**	**1.86**
HADS-A	7.47 (4.07)	7.43 (4.09)	7.64 (3.99)	0.21
HADS-D	5.88 (4.08)	5.76 (3.98)	6.40 (4.44)	0.64
Pain: least	4.39 (2.76)	4.34 (2.78)	4.58 (2.67)	0.24
Pain: Average last two weeks	6.62 (2.37)	6.55 (2.41)	6.88 (2.18)	0.33
Pain: Current	4.91 (2.63)	4.81 (2.62)	5.34 (2.64)	0.53

Responders at four-month follow-up were older, less disabled (RMDQ and SF-12) and catastrophised less. Smaller differences between other measures but indication of responders being less fear-avoidant, less anxious, less depressed and reporting less pain.

High-risk Group only

Measure	**Baseline values (mean (SD)) (*n*=851)**	**Baseline values (Mean (SD)) at four-month follow-up**		
		Resp (*n*=188)	**Non-resp (*n*=48)**	***Difference***
Age	**49.73 (14.76)**	**53.95 (14.13)**	**43.33 (14.39)**	**10.62**
Sex	500 (47%) f	111 (59%) f	22 (46%) f	
RMDQ	**9.76 (5.67)**	**14.28 (4.56)**	**15.40 (4.46)**	**1.12**
SF-12 PCS	**36.80 (10.47)**	**31.12 (7.84)**	**27.61 (7.19)**	**3.51**
TSK	**40.42 (6.20)**	**45.49 (5.19)**	**47.30 (5.25)**	**1.81**
PCS	**16.14 (10.86)**	**20.04 (10.25)**	**28.67 (11.27)**	**8.63**
HADS-A	7.47 (4.07)	10.23 (3.96)	9.60 (4.25)	0.63
HADS-D	5.88 (4.08)	8.73 (4.02)	9.56 (4.42)	0.83
Pain: least	4.39 (2.76)	6.21 (2.73)	5.96 (2.75)	0.25
Pain: Average last two weeks	6.62 (2.37)	7.86 (1.92)	8.44 (1.62)	0.58
Pain: Current	4.91 (2.63)	6.56 (2.32)	7.00 (1.98)	0.44

Much larger differences in the high-risk group – responders at four-month follow-up are older, less disabled (RMDQ and SF-12), catastrophised less and were less fear-avoidant. Smaller differences between other measures but responders were more anxious and less depressed.

High-risk Group only: Split by treatment and control

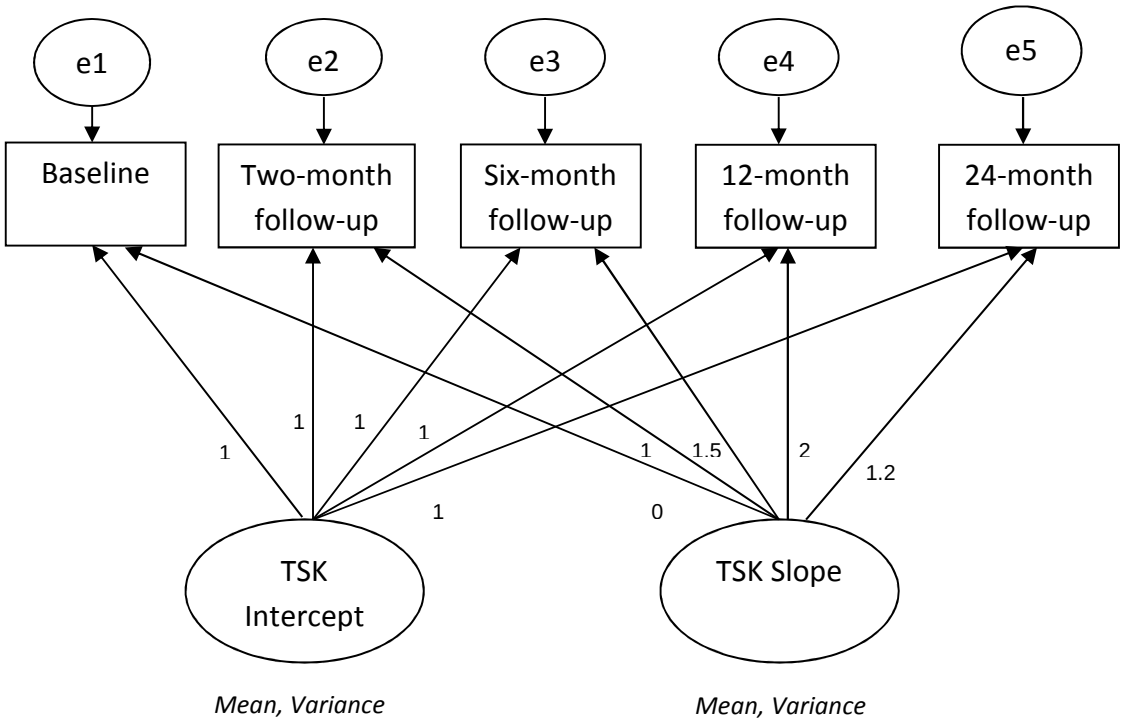
Measure	**Baseline values (mean (SD)) (*n*=851)**	**Baseline values (Mean (SD)) at four-month follow-up**					
		Treatment		**Control**			
		Resp (*n*=131)	**Non-resp (*n*=26)**	**Resp (*n*=57)**	**Non-resp (*n*=22)**	***Difference***	
						Treatment	***Control***
Age	**49.73 (14.76)**	**53.87 (13.85)**	**46.50 (16.42)**	**54.12 (14.88)**	**39.59 (10.74)**	**7.37**	**14.53**
Sex	500 (47%) f	76 (58%) f	12 (46%) f	35 (61%) f	10 (46%) f		
RMDQ	**9.76 (5.67)**	**14.27 (4.54)**	**15.08 (5.15)**	**14.32 (4.65)**	**15.77 (3.56)**	**0.81**	**1.45**
SF-12 PCS	**36.80 (10.47)**	**31.37 (7.67)**	**28.13 (6.72)**	**30.56 (8.25)**	**27.00 (7.83)**	**3.24**	**3.56**
TSK	**40.42 (6.20)**	**45.57 (5.12)**	**47.05 (4.48)**	**45.32 (5.38)**	**47.60 (6.13)**	**1.48**	**2.28**
PCS	**16.14 (10.86)**	**25.86 (10.32)**	**29.19 (11.89)**	**26.45 (10.15)**	**28.05 (10.73)**	**3.33**	**1.60**
HADS-A	7.47 (4.07)	10.17 (4.18)	9.81 (4.08)	10.37 (3.44)	9.36 (4.53)	0.36	**1.01**
HADS-D	5.88 (4.08)	8.82 (4.20)	9.27 (4.94)	8.53 (3.60)	9.91 (3.80)	0.45	**1.38**
Pain: least	4.39 (2.76)	6.32 (2.55)	6.38 (2.62)	5.98 (3.12)	5.45 (2.87)	0.06	0.53
Pain: Average last two weeks	**6.62 (2.37)**	**7.78 (1.97)**	**8.69 (5.12)**	**8.02 (1.79)**	**8.14 (1.73)**	**0.91**	0.12
Pain: Current	**4.91 (2.63)**	**6.56 (2.27)**	**7.27 (1.76)**	**6.54 (2.49)**	**6.68 (2.21)**	**0.71**	0.14

In the treatment group, patients responding at four-month follow-up were older, less disabled (RMDQ and SF-12), catastrophised less and were less fear-avoidant than those who did not respond. They also reported less average and current pain.

In the control group, patients responding at four-month follow-up were older, less disabled (RMDQ and SF-12), less fear-avoidant and catastrophised less, and also reported more anxiety and less depression. No differences in pain between responders and non-responders.

Appendix 7.1: Sensitivity analysis for Back In Action study: Complete Case analysis (based on responses to RMDQ and TSK) (*n*=172)

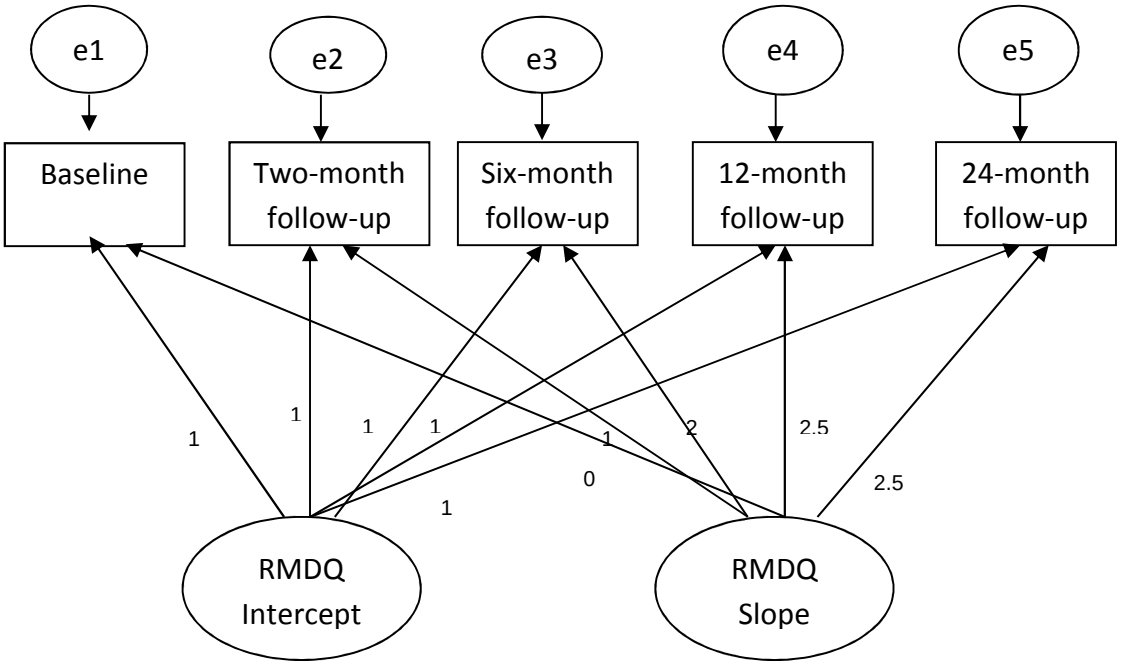
LGM Model for Fear-avoidance beliefs (10-item TSK)



Model Fit Statistics for Growth Curve Model of 10-item TSK

Model Index	**Current Model**	**Good Model Fit**
CMIN (Chi Square)	46.91	Non-significant result
DF	14	
P	0.00	
CMIN/DF	3.35	Between 2-5
CFI	0.95	>0.95
RMSEA	0.12	<0.08
SRMR	0.04	<0.08

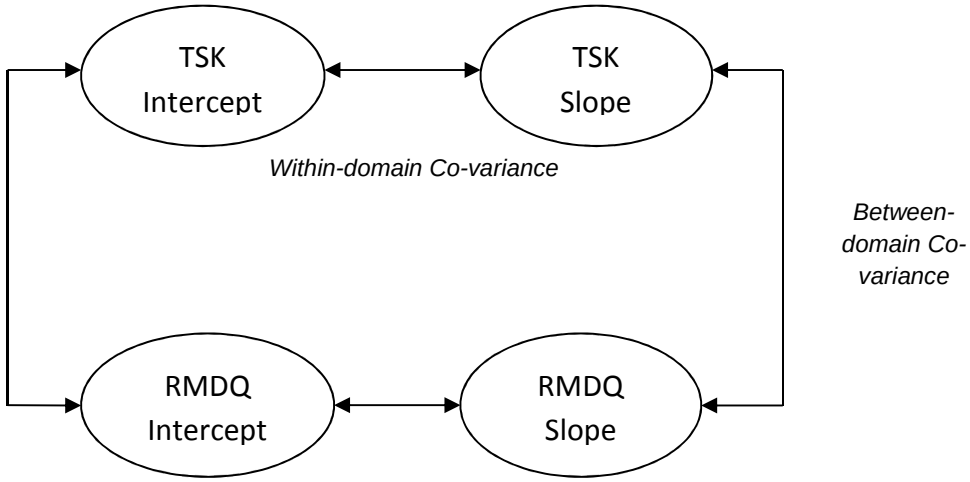
LGM Model for Disability (23-item RMDQ)



Model Fit Statistics for Growth Curve Model of RMDQ

Model Index	**Current Model**	**Good Model Fit**
CMIN (Chi Square)	26.27	Non-significant result
DF	15	
P	0.04	
CMIN/DF	1.75	Between 2-5
CFI	0.98	>0.95
RMSEA	0.07	<0.08
SRMR	0.03	<0.08

Simplified LGM Model to show Relationships between Mediator and Outcome Variables



Model Fit Statistics for Parallel Process Model without Treatment Group Allocation as a Co-variate

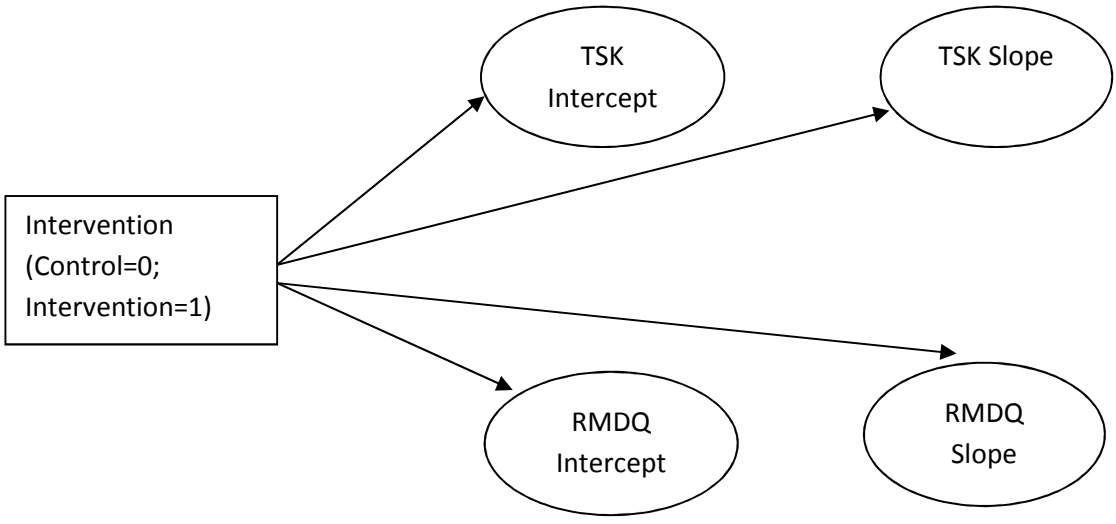
Model Index	**Original Model**	**Modified Model**	**Good Model Fit**
CMIN* (Chi Square)	152.62	135.74	Non-significant result
DF	51	50	
P	0.00	0.00	
CMIN/DF	2.99	2.72	Between 2-5
CFI	0.93	0.94	>0.95
RMSEA	0.11	0.10	<0.08
SRMR	0.04	0.04	<0.08

Table 7.7 Means, Co-variances, Correlations and Variances between Mediator and Outcome Variables

	Variables	**Estimate**	**Standard Error**	***p*-value**
Means	TSK Intercept (*baseline value*)	40.42	0.64	<0.05
	TSK Slope (*change over all time points*)	-2.84	0.32	<0.05
	RMDQ Intercept	11.55	0.42	<0.05
	RMDQ Slope	-1.18	0.15	<0.05
Co-variances	TSK Intercept <-> TSK Slope	1.64	2.53	0.52
	TSK Intercept <-> RMDQ Intercept	22.50	3.76	<0.05
	TSK Slope <-> RMDQ Slope	2.98	0.62	<0.05
	RMDQ Intercept <-> RMDQ Slope	0.57	0.74	0.44
Variances	TSK Intercept	52.41	7.56	<0.05
	TSK Slope	7.30	2.02	<0.05
	RMDQ Intercept	23.35	3.30	<0.05
	RMDQ Slope	1.52	0.41	<0.05

β = Unstandardised estimate
b = Standardised estimate

Simplified LGM Model to show Relationships between Mediator and Outcome Variables, with Treatment Group Allocation as a Co-variate



Model Fit Statistics for Parallel Process Model with Treatment Group Allocation Co-variate

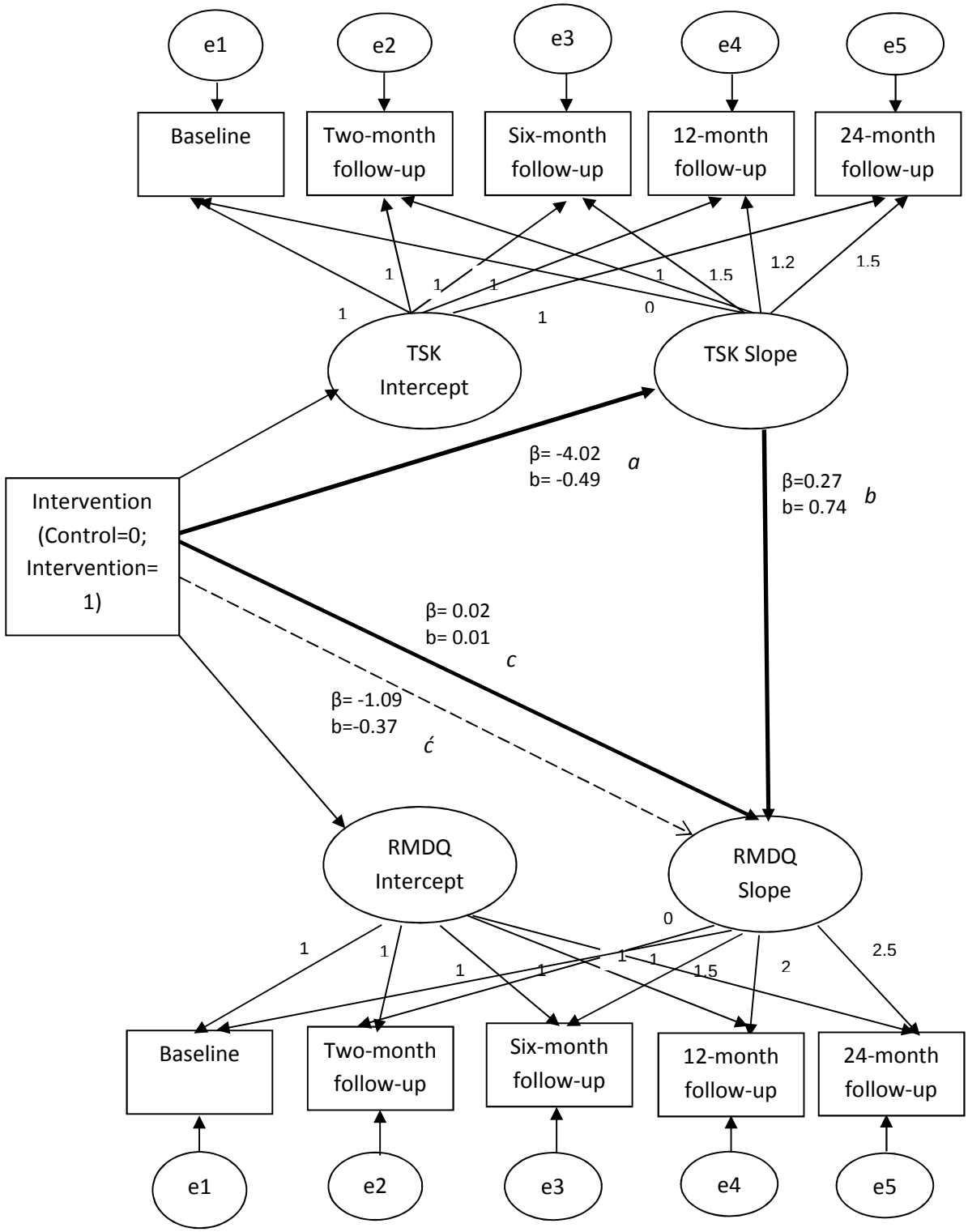
Model Index	**Original Model**	**Modified Model**	**Good Model Fit**
CMIN (Chi Square)	263.49	168.18	Non-significant result
DF	61	59	
P	0.00	0.00	
CMIN/DF	4.31	2.85	Between 2-5
CFI	0.86	0.93	>0.95
RMSEA	0.14 (0.12 to .16, PCLOSE 0.00)	0.10 (0.09 to 0.12, PCLOSE 0.00)	<0.08
SRMR	0.27	0.04	<0.08

Mean Values for Mediator and Outcome, Related to Treatment Group Allocation

	Variables	**Estimate**	**Standard Error**	**p-value**
Regression Coefficients	TSK Intercept	β -0.94 b 0.06	1.30	0.47
	TSK Slope	β -2.18 b -0.40	0.63	<0.05
	RMDQ Intercept	β 0.94 b 0.10	0.86	0.27
	RMDQ Slope	β -0.93 -0.34	0.30	0.00

β = Unstandardised estimate
b = Standardised estimate

Full Mediation Model: Change in Fear-avoidance beliefs (10-item TSK) as a Mediator between Intervention Allocation and Change in Disability Outcome (23-item RMDQ)



β = Unstandardised estimate
b = Standardised estimate

Mediating Effect of Change in Fear-avoidance beliefs (TSK) on the Relationship between Treatment Allocation and Change in Disability (RMDQ)

	Effect	Model	
		Standardised Estimates (95% CI)	**Unstandardised Estimates (95% CI)**
RMDQ (23-item)$^{\Delta}$	Total (*ć*)	-0.37 (-0.54 to -0.14)	-1.09 (-1.69 to -0.40)
	Direct (*c*)	0.01 (-0.27 to 0.27)	0.02 (-0.76 to 0.80)
	Indirect (*ab*)	-0.37 (-0.54 to -0.22)	-1.10 (-1.73 to -0.59)

Model Fit Statistics for Full Mediation Model

Model Index	**Full Mediation Model**	**Modified Model**	**Good Model Fit**
CMIN (Chi Square)	811.92	295.98	Non-significant result
DF	63	59	
P	0.00	0.00	
CMIN/DF	12.89	5.02	Between 2-5
CFI	0.50	0.84	>0.95
RMSEA	0.26	0.15	<0.08
SRMR	0.36	0.25 (0.14 to 0.17, PCLOSE 0.00)	<0.08

Appendix 7.2: Assessing Trajectories of RMDQ and TSK – Model Fit

RMDQ: Treatment Trajectory (Both groups)

Factor Score Sequence	CMIN (df)	CFI	RMSEA	SRMR
0, -0.2, -0.5, -1, -2		0.92	0.14	0.03
0, -1, -1, -1.5, -3		0.91	0.15	0.05
0, -1, -1.5, -2, -3		0.95	0.11	0.04
0, -0.2, -0.4, -1, -2		0.91	0.15	0.03
0, -1, -1.5, -2, -2.5		**0.96**	**0.10**	**0.04**

RMDQ: Treatment Trajectory (Control group only)

CFI = 0.93, RMSEA = 0.13, SRMR = 0.05

RMDQ: Control Trajectory

0, 0, -1, -2, -2: CFI = 0.97, RMSEA = 0.08, SRMR = 0.03

TSK: Treatment Trajectory (Both groups)

Factor Score Sequence	CMIN (df)	CFI	RMSEA	SRMR
0, -0.2, -0.5, -1, -2		0.85	0.20	0.05
0, -1.2, -1.5, -1, -2		0.90	0.17	0.04
0, -1.2, -1.5, -1, -1.5		0.91	0.15	0.04
0, -1.2, -1.5, -1.2, -2.5		0.89	0.17	0.04
0, -1.2, -1.5, -1, -2.5		0.87	0.18	0.05
0, -1.2, -1.5, -1.2, -1.5		0.94	0.13	0.04
0, -1, -1.5, -1.2, -1.5		**0.95**	**0.12**	**0.04**
0, -1, -1.4, -1, -1.7		0.92	0.15	0.04
0, -1, -1.5, -1.2, -1.6		0.94	0.12	0.04

TSK: Treatment Trajectory (Control group only)

CFI = 0.91, RMSEA = 0.15, SRMR = 0.06

TSK: Control Trajectory

0, -0.5, -1, -2, -1: CFI = 0.94, RMSEA = 0.12, SRMR = 0.05

Appendix 8.1 Distribution of SIP Disability Score

	Mean (SD)	Skewness (SE)	Kurtosis (SE)
Post-treatment SIP Disability score	0.16 (0.14)	1.02 (0.29)	1.27 (0.54)
Residualised change in SIP Disability (pre-post treatment)	0.00 (1.00)	0.24 (0.29)	0.32 (0.57)